

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 16:01:26 ; Search time 5499 Seconds
(without alignments)
11302.766 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaactcttttctcg.....aaaaaaaaaaaaaaaaa 1434

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	GenEmbl :
1: gb_ba.*	BD078422
2: gb_hgt.*	101 human secretory proteins.
3: gb_in.*	BD078422
4: gb_om.*	BD078422
5: gb_ov.*	JP 2001519156-A/11
6: gb_pat.*	KEYWORDS Homo sapiens (human)
7: gb_ph.*	SOURCE Homo sapiens
8: gb_pl.*	ORGANISM Homo sapiens
9: gb_pr.*	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
10: gb_ro.*	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
11: gb_sts.*	Duan,R.D., Florence,K.A., Rosen,C.A., Ruben,S.M., Greene,J.M.,
12: gb_sy.*	Young,P., Ferrie,A.M., Yu,G.L., Janat,P.Ni,J., Carter,K.C.,
13: gb_un.*	Endress,G.A., Feng,P., Lafleur,D.W. and Shi,Y.
14: gb_vi.*	
15: em_ba.*	
16: em_fun.*	
17: em_hum.*	
18: em_in.*	
19: em_mu.*	
20: em_om.*	
21: em_or.*	
22: em_ov.*	
23: em_pat.*	
24: em_ph.*	
25: em_pl.*	
26: em_ro.*	
27: em_sts.*	
28: em_un.*	
29: em_vi.*	
30: em_hgt_hum.*	
31: em_hgt_inv.*	
32: em_hgt_other.*	
33: em_hgt_mus.*	
34: em_hgt_pln.*	
35: em_hgt_rod.*	
36: em_hgt_man.*	
37: em_hgt_vrt.*	
38: em_sy.*	
39: em_hgtgo_hum.*	
40: em_hgtgo_mus.*	
41: em_hgtgo_other.*	

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1432	99.9	1434	6	BD078422	BD078422 101 human
2	1394.8	97.3	3522	9	HSM804706	AL833393 Homo sapi
3	1373.8	95.8	2020	9	HSM800227	AL049442 Homo sapi
4	1368.4	95.4	3452	9	BC033650	BC033650 Homo sapi
c 5	1359.8	94.8	175081	9	AC021151	AC021151 Homo sapi
c 6	591	41.2	281662	2	AC117374	AC117374 Homo sapi
c 7	555	38.7	281662	2	AC117374	AC117374 Homo sapi
c 8	425.8	29.7	448	6	AX408738	AX408738 Sequence
c 9	356.8	24.9	395	6	AX898296	AX898296 Sequence
c 10	356.8	24.9	395	6	BD033829	BD033829 Sequence
c 11	84.8	5.9	184359	10	AC121881	AC121881 Mus muscu
c 12	82	5.7	210911	2	AC115265	AC115265 Rattus no
c 13	77.4	5.4	212827	9	AC009412	AC009412 Homo sapi
c 14	76.6	5.3	72832	2	AC099821	AC099821 Homo sapi
c 15	73.2	5.1	110000	3	AC116984_1	Continuation (2 of
c 16	72.2	5.0	2465	3	AF238313	AF238313 Dictyoste
c 17	70.6	4.9	2192	3	AF298207	AF298207 Dictyoste
c 18	70.6	4.9	192929	2	AC005505	AC005505 Plasmodiu
c 19	70.6	4.9	250713	3	AE014850	AE014850 Plasmodiu
c 20	69.8	4.9	545	6	AX185705	AX185705 Sequence
c 21	69.6	4.9	1357	5	BC054578	BC054578 Danio rer
c 22	69.2	4.8	5325	3	AY160094	AY160094 Dictyoste
c 23	69	4.8	8056	6	AX599046	AX599046 Sequence
c 24	68.8	4.8	110000	3	PFMAL1F2_1	Continuation (2 of
c 25	68.4	4.8	110000	3	AC116984_1	Continuation (2 of
c 26	68.2	4.8	64707	3	AC115607	AC115607 Dictyoste
c 27	68.2	4.8	254436	3	AE014827	AE014827 Plasmodiu
c 28	68	4.7	189218	2	EX897692	EX897692 Danio rer
c 29	67.8	4.7	302156	3	AC116977	AC116977 Dictyoste
c 30	67.6	4.7	125623	3	AC115599	AC115599 Dictyoste
c 31	67.6	4.7	182870	3	AC116960	AC116960 Dictyoste
c 32	67.6	4.7	266544	3	AC116956	AC116956 Dictyoste
c 33	67.4	4.7	125958	3	AC115592	AC115592 Dictyoste
c 34	67	4.7	110000	2	PFMAL8P1_00	AL844507 Plasmodiu
c 35	67	4.7	110000	2	PFMAL8P1_01	Continuation (2 of
c 36	66.6	4.6	1092	3	DDSTATCUL	Y13098 Dictyosteli
c 37	66.6	4.6	254436	3	AE014827	AE014827 Plasmodiu
c 38	66.2	4.6	110000	2	PFMAL7P1_03	Continuation (4 of
c 39	66	4.6	166447	2	BX323824	BX323824 Danio rer
c 40	65.6	4.6	597	6	AX187052	AX187052 Sequence
c 41	65.6	4.6	2067	3	FFA271108	AX271108 Plasmodiu
c 42	65.6	4.6	254733	3	AC117075	AC117075 Dictyoste
c 43	65.4	4.6	110000	2	PFMAL13_19	Continuation (20 o
c 44	65.4	4.6	136240	3	AC117070	AC117070 Dictyoste
c 45	65.2	4.5	41399	3	AC116984_5	Continuation (6 of

ALIGNMENTS

RESULT 1	LOCUS	BD078422	101 human secretory proteins.	1434 bp	DNA	linear	PAT 27-AUG-2002
BD078422	DEFINITION	BD078422	101 human secretory proteins.				
BD078422	ACCESSION	BD078422	101 human secretory proteins.				
BD078422	VERSION	BD078422.1	GI:22624025				
BD078422	KEYWORDS	JP 2001519156-A/11					
BD078422	SOURCE	JP 2001519156-A/11					
BD078422	ORGANISM	Homo sapiens					
BD078422	REFERENCE	1 (bases 1 to 1434)					
BD078422	AUTHORS	Duan,R.D., Florence,K.A., Rosen,C.A., Ruben,S.M., Greene,J.M.,					
BD078422		Young,P., Ferrie,A.M., Yu,G.L., Janat,P.Ni,J., Carter,K.C.,					
BD078422		Endress,G.A., Feng,P., Lafleur,D.W. and Shi,Y.					

TITLE 101 human secretory proteins
 JOURNAL Patent: JP 2001519156-A 11 23-OCT-2001;
 HUMAN GENOME SCIENCES INC
 OS Homo sapiens (human)
 PN JP 2001519156-A/11
 PD 23-OCT-2001
 PF 01-OCT-1998 JP 2000515006
 PR 02-OCT-1997 US 60/060837, 02-OCT-1997 US 60/060862 PR
 02-OCT-1997 US 60/060839, 02-OCT-1997 US 60/060866 PR
 02-OCT-1997 US 60/060843, 02-OCT-1997 US 60/060836 PR
 02-OCT-1997 US 60/060838, 02-OCT-1997 US 60/060874 PR
 02-OCT-1997 US 60/060833, 02-OCT-1997 US 60/060884 PR
 02-OCT-1997 US 60/060880
 PI ROXANNE D DUAN, KIMBERLY A FLORENCE, CRAIG A ROSEN, STEVEN M PI
 RUBEN,
 PI JOHN M GRENE, PAUL YOUNG, ANN M FERRIE, GUO
 LIANG YU, FOJAD JANAT,
 PI JIAN NI,
 PI KENNETH C CARTER, GREGORY A ENDRESS, PING FENG, DAVID W LAFLEUR,
 PI YANGGU SHI
 PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P25/00, A61P35/00,
 PC A61P37/00,
 PC A61P43/00, C07K14/47, C07K16/00, C12N5/10, C12P21/02, C12Q1/68, PC
 G01N33/53,
 PC G01N33/53, G01N33/566, C12N15/00, A61K37/02, C12N5/00 CC 101
 human secretory proteins
 FH Key Location/Qualifiers
 FT source 1..1434
 FT /organism="Homo sapiens (human)".
 FT /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

Query Match 99.9%; Score 1432; DB 6; Length 1434;
 Best Local Similarity 100.0%; Pred. No. 2.3e-227;
 Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATTAACTCTTTTATCGGAATAGTATGATATTTCAATGTCACCTCCCATTCATGTTGA 60
 DB 1 CATTAACTCTTTTATCGGAATAGTATGATATTTCAATGTCACCTCCCATTCATGTTGA 60
 QY 61 TTGGAGCTCACAGTTATTTTGTGTAAGCAGAGATTTAAATTTATTTGTAAGTCAGTGC 120
 DB 61 TTGGAGCTCACAGTTATTTTGTGTAAGCAGAGATTTAAATTTATTTGTAAGTCAGTGC 120
 QY 121 AAAATTATGAATAGGATATATACTAATAATCAAAAGTAAATAACAAAGTCAAAAGCAGTGT 180
 DB 121 AAAATTATGAATAGGATATATACTAATAATCAAAAGTAAATAACAAAGTCAAAAGCAGTGT 180
 QY 181 CTAATAAATAATCTGGGTTCCCTTAAATAATTTTAAATTTATCTTGAATAGTTTCT 240
 DB 181 CTAATAAATAATCTGGGTTCCCTTAAATAATTTTAAATTTATCTTGAATAGTTTCT 240
 QY 241 TAGAATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300
 DB 241 TAGAATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300
 QY 301 CAAATTGCTATTAAATGCMAGAGTGGAATATATACAGAAATTTATCAGGCATTACCAAGTC 360
 DB 301 CAAATTGCTATTAAATGCMAGAGTGGAATATATACAGAAATTTATCAGGCATTACCAAGTC 360
 QY 361 TAGGCATATAGGAATGAGCACTCAGAAATGTTTCAATGATAGTTGATGCTTGTGA 420
 DB 361 TAGGCATATAGGAATGAGCACTCAGAAATGTTTCAATGATAGTTGATGCTTGTGA 420
 QY 421 AGGTAGGGAGCTTATTACAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATTTGCTGG 480
 DB 421 AGGTAGGGAGCTTATTACAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATTTGCTGG 480
 QY 481 CCTTTGGCTACCTGTACTTCCSCAATATGGAGCCCAATTCAGTCTTCTTCTCT 540

DB 481 CCTTTGGCTACCTGTACTTCCSCAATATGGAGCCCAATTCAGTCTTCTTCTCT 540
 QY 541 GGCACACTTATGCTCTGAAATCATGAGCGAGGCTGATTCATTTGGTGGTGGTAGAAA 600
 DB 541 GGCACACTTATGCTCTGAAATCATGAGCGAGGCTGATTCATTTGGTGGTGGTAGAAA 600
 QY 601 GCAGTATGTTTGTGTCATTAAGATGTAGGTATATAGATAGTTTACGCTTTAAGTGAT 660
 DB 601 GCAGTATGTTTGTGTCATTAAGATGTAGGTATATAGATAGTTTACGCTTTAAGTGAT 660
 QY 661 GTTTTATATCTTTAAATAAAGAAATATAACCTTTTAAAGCTATTCCACCTCTCTCCCCAGC 720
 DB 661 GTTTTATATCTTTAAATAAAGAAATATAACCTTTTAAAGCTATTCCACCTCTCTCCCCAGC 720
 QY 721 CTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCACT 780
 DB 721 CTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCACT 780
 QY 781 GCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTTAATAC 840
 DB 781 GCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTTAATAC 840
 QY 841 TTACTTCTATTTCTGAAATCAGTTGTGAACCTGTTGCTTTATGTTCAAGAGTTTAAAGAAC 900
 DB 841 TTACTTCTATTTCTGAAATCAGTTGTGAACCTGTTGCTTTATGTTCAAGAGTTTAAAGAAC 900
 QY 901 CTCMGTAATTCATTTTAAATCTGCTATTCGAGAAGCAATGTAATGTAATTTTCAACA 960
 DB 901 CTCMGTAATTCATTTTAAATCTGCTATTCGAGAAGCAATGTAATGTAATTTTCAACA 960
 QY 961 AGAAGCTCATCTGAGTCTTTTCTGCTGACCTCTATGAGCCCATAGGGTCTGCTGCTTA 1020
 DB 961 AGAAGCTCATCTGAGTCTTTTCTGCTGACCTCTATGAGCCCATAGGGTCTGCTGCTTA 1020
 QY 1021 GCATTAACAAATAAGGTTTATAGTAAAGCCCAATGTAATTTTTCATGTCAGG 1080
 DB 1021 GCATTAACAAATAAGGTTTATAGTAAAGCCCAATGTAATTTTTCATGTCAGG 1080
 QY 1081 GCTTTAAATTTGCTCTCTTTTCAATTTTATTCATATTCATTTTATGCTTTGTAAGT 1140
 DB 1081 GCTTTAAATTTGCTCTCTTTTCAATTTTATTCATATTCATTTTATGCTTTGTAAGT 1140
 QY 1141 CTTTATAGGAGATAATATATGTTTAAATTTAGTTTGGGGGGAATAATGTCGAAGA 1200
 DB 1141 CTTTATAGGAGATAATATATGTTTAAATTTAGTTTGGGGGGAATAATGTCGAAGA 1200
 QY 1201 GGATAATTTAATTTACGTGCTTCTGTTTATTCAGAAATAAGAGAGAGACTACGTCGATA 1260
 DB 1201 GGATAATTTAATTTACGTGCTTCTGTTTATTCAGAAATAAGAGAGAGACTACGTCGATA 1260
 QY 1261 TTCAAGAGTTGTACCTTAACATTTGTTGTAACATTTTCTAAGATTTTCAAAAGGAATAT 1320
 DB 1261 TTCAAGAGTTGTACCTTAACATTTGTTGTAACATTTTCTAAGATTTTCAAAAGGAATAT 1320
 QY 1321 GTCTAAATTCAGAAATCATACCACTGCTCTAACTTGGTAAACAACTGTTCTTAATAA 1380
 DB 1321 GTCTAAATTCAGAAATCATACCACTGCTCTAACTTGGTAAACAACTGTTCTTAATAA 1380
 QY 1381 AGTATTTAATGATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1434
 DB 1381 AGTATTTAATGATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1434

RESULT 2
 HSM804706
 LOCUS Homo sapiens mRNA; cDNA DKFp762K109 (from clone DKFp762K109).
 DEFINITION
 AL833393
 ACCESSION
 VERSION
 AL833393.1 GI:21734029
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3522)
Ostenwaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and Wiemann, S.
Direct Submission
Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MediGenomix (Martinried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp762K109) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clonezrpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES
source
1..3522
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="RZPD:DKFZp762K109"
 /db_xref="taxon:9606"
 /clone="DKFZp762K109"
 /tissue_type="melanoma (MeWo cell line)"
 /clone_lib="762 (synonym: hmel2). Vector pSport1; host DH10B; sites NotI + SalI"
 /dev_stage="adult"
polyA_signal
 3420..3425
polyA_site
 3440
ORIGIN

Query Match 97.3%; Score 1394.8; DB 9; Length 3522;
Best Local Similarity 99.3%; Pred. No. 2.8e-221;
Matches 1426; Conservative 5; Mismatches 22; Indels 3; Gaps 3;

QY 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTACACCTCCATTCATGTGA 60
DB 2044 CATTAACCTCTTTTATCGGGAATAGTATGATATTTTCAATGTACACCTCCATTCATGTGA 2103

QY 61 TTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTTAAATTTTATATGAAAGTCAGTGC 120
DB 2104 TTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTTAAATTTTATATGAAAGTCAGTGC 2163

QY 121 AAAATTATGAATAGGATATCTAATAATAACAAGTAATTAACAAGTCAAGAGTGT 180
DB 2164 AAAATTATGAATAGGATATCTAATAATAACAAGTAATTAACAAGTCAAGAGTGT 2223

QY 181 CTAATAAAAAATTCGGTTCCTTAAATAATTTTAAATTTTATC-TTGAATAGTTTTC 239
DB 2224 CTAATAAAAAATTCGGTTCCTTAAATAATTTTAAATTTTATC-TTGAATAGTTTTC 2283

QY 240 TTAGATTAATCTCAGGATATGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 299
DB 2284 TTAGATTAATCTCAGGATATGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 2343

QY 300 ACAAATTGCTATTAATGACAGA-GTGGTAATATACAGAAATTAACAGGCATACCAAG 358
DB 2344 ACAAATTGCTATTAATGACAGCGTGGTAATATACAGAAATTAACAGGCATTAACCAAG 2403

QY 359 TCTAGGCACATATAGGAATGCGAGCACTCAGATGGTTTCAATGTAGTAGTTCATGCTTG 418
DB 2404 TCTAGGCACATATAGGAATGCGAGCACTCAGATGGTTTCAATGTAGTAGTTCATGCTTG 2463

QY 419 TAAGGTAGGGAGCTTATTCAGACATAGTAGTAGTTTCTTAATGCTGTSTCAATTTGCT 478
DB 2464 TAAGGTAGGGAGCTTATTCAGACATAGTAGTAGTTTCTTAATGCTGTSTCAATTTGCT 2523

QY 479 GGCCTTTGGCTACTGCTACTTCCTCCATTATGGCAGCCCATTCAGTCTTGATTTCTTCT 538
DB 2524 GGCCTTTGGCTACTGCTACTTCCTCCATTATGGCAGCCCATTCAGTCTTGATTTCTTCT 2583

QY 539 CTGGACACCTTATGCTCTGAATATCATGAGGAGGCTGATTCATTTGGTGAATTTGGGTAGA 598
DB 2584 CTGGACACCTTATGCTCTGAATATCATGAGGAGGCTGATTCATTTGGTGAATTTGGGTAGA 2643

QY 599 AAGCAGTATCTTTTGTGACATTAAGATGTAGGTATAGATAGTCTTTAGCCTTTAAGTGT 658
DB 2644 AAGCAGTATCTTTTGTGACATTAAGATGTAGGTATAGATAGTCTTTAAGTGT 2703

QY 659 ATGTTTTTATCTTTTAAATAAGAAATATAACCTTTTAACTATTCACCTCTCTCCCCCA 718
DB 2704 ATGTTTTTATCTTTTAAATAAGAAATATAACCTTTTAACTATTCACCTCTCTCCCCCA 2763

QY 719 GCCTATCTCAAACTGGTGGAAATATATGGAGAGATCTTTGAAAGAGTAAATAAACCCTTCA 778
DB 2764 GCCTATCTCAAACTGGTGGAAATATATGGAGAGATCTTTGAAAGAGTAAATAAACCCTTCA 2823

QY 779 CTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTATTTAAT 838
DB 2824 CTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTATTTAAT 2883

QY 839 ACTTACCTTCTATTTCTGAATCAGTTGTGAACCTGTTCCTTATGTTTCAGAGTTTAAAGA 898
DB 2884 ACTTACCTTCTATTTCTGAATCAGTTGTGAACCTGTTCCTTATGTTTCAGAGTTTAAAGA 2943

QY 899 ACCTCMGTGAATTCATTTTAAATCTGCTATTCTGAGAGCATTGAATGAATCTTAA 958
DB 2944 ACCTCMGTGAATTCATTTTAAATCTGCTATTCTGAGAGCATTGAATGAATCTTAA 3003

QY 959 CAAGAAGACTCATCTGTAGCTGTTTGTGATCTCTATGAGCCCAATAAGGTTCTGTGCT 1018
DB 3004 CAAGAAGACTCATCTGTAGCTGTTTGTGATCTCTATGAG-CCCAATAAGGTTCTGTGCT 3062

QY 1019 TACATTAACAAAATAAGTTTATAGTAAAGCAATGATTAATTTTTTTTTCATGGA 1078
DB 3063 TAGCATTAACAAAATAAGTTTATAGTAAAGCAATGATTAATTTTTTTTTCATGGA 3122

QY 1079 GGGCTTTAAATTTGTGCTCTTTTCAATATTTTATCATATTCATATTTATGTTTGAAC 1138
DB 3123 GGGCTTTAAATTTGTGCTCTTTTCAATATTTTATCATATTCATATTTATGTTTGAAC 3182

QY 1139 TGCTTTTATAGGAGATAATTAATGTTTAAATTTAGTTTGGGGGAATAATTTGTGCAA 1198
DB 3183 TGCTTTTATAGGAGATAATTAATGTTTAAATTTAGTTTGGGGGAATAATTTGTGCAA 3242

QY 1199 GAGATAATTTAATTTACGTGCTCTGTTTATTCAGATAAAGAGAGAAGCTACGCTGCA 1258
DB 3243 GAGATAATTTAATTTACGTGCTCTGTTTATTCAGATAAAGAGAGAAGCTACGCTGCA 3302

QY 1259 TATTCAAGAGTTGCTACCTTAACATTCGTGAAACATTTTTTCTAAGATTTTCAAAAGAAAT 1318
DB 3303 TATTCAAGAGTTGCTACCTTAACATTCGTGAAACATTTTTTCTAAGATTTTCAAAAGAAAT 3362

QY 1319 ATGTGTAATTTGAGAAATCATACCACTGCTCTTAACCTGGTAACAACTGTTCTTAAAT 1378
DB 3363 ATGTGTAATTTGAGAAATCATACCACTGCTCTTAACCTGGTAACAACTGTTCTTAAAT 3422

QY 1379 AAGTATTTAATGATTTTAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1434
DB 3423 AAGTATTTAATGATTTTAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 3478

RESULT 3
HSM800227
LOCUS
DEFINITION
Homo sapiens mRNA; cDNA DKFZp586N1720 (from clone DKFZp586N1720).
ACCESSION
AL049442
VERSION
AL049442.1 GI:4500222
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2020)
AUTHORS
Wambutt, R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE
Direct Submission
JOURNAL
Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152

COMMENT
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOMA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp586N1720) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES
Location/Qualifiers
1..2020
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp586N1720"
/db_xref="taxon:9606"
/clone="DKFZp586N1720"
/tissue type="uterus"
/clone_lib="586 (synonym: hutel). Vector pSport1; host DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
polyA_signal
1983..1988
polyA_site
2003
ORIGIN

Query Match 95.8%; Score 1373.8; DB 9; Length 2020;
Best Local Similarity 95.3%; Pred. No. 9.2e-218; Mismatches 5; Gaps 3;
Matches 1405; Conservative 2; Indels 3; Gaps 3;

Qy 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTCACTCCCATTCATGTTGA 60
Db 607 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTCACTCCCATTCATGTTGA 666

Qy 61 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTAAATTTATATGTAAGTCAGTGC 120
Db 667 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTAAATTTATATGTAAGTCAGTGC 726

Qy 121 AAAATTATGAATAGGATATCTAATAATACAAAGTAAATACAAAGTCAAAAGCAGGTT 180
Db 727 AAAATTATGAATAGGATATCTAATAATACAAAGTAAATACAAAGTCAAAAGCAGGTT 786

Qy 181 CTAATAAAAATCTCGGTTCTTTTAAATTTATTTTAAATTTATC-TTGAATAGTTTTC 239
Db 787 CTAATAAAAATCTCGGTTCTTTTAAATTTATTTTAAATTTATC-TTGAATAGTTTTC 846

Qy 240 TTAGATTAACTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 299
Db 847 TTAGATTAACTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 906

Qy 300 ACAAATTGTCTATTAAATGCMGA-GTGGTAAATACAGAAATTTATCAGGCATTACCAAG 358
Db 907 ACAAATTGTCTATTAAATGCMGACGTGGTAAATACAGAAATTTATCAGGCATTACCAAG 966

Qy 359 TCTAGGCACATATAGGAATGACGACCTCAGAAATGGTTTCAATGATAGTATGATGCTTG 418
Db 967 TCTAGGCACATATAGGAATGACGACCTCAGAAATGGTTTCAATGATAGTATGATGCTTG 1026

Qy 419 TAAGGTAGGGAGCTTATTCAGACATAGTATGATTTCTCTAATGCTGTCTCAATTCCT 478
Db 1027 TAAGGTAGGGAGCTTATTCAGACATAGTATGATTTCTCTAATGCTGTCTCAATTCCT 1086

Qy 479 GGCCTTTGGGTACCTGTACTTCCSCATTATGGCAGCCCATTCAGTCTTGAGTTTCTTCT 538
Db 1087 GGCCTTTGGGTACCTGTACTTCCSCATTATGGCAGCCCATTCAGTCTTGAGTTTCTTCT 1146

Qy 539 CTGACACCTTATGCTCTGAAATCATGAGCAGGCTGATTCATTTGGTGAATTTGGGTAGA 598
Db 1147 CTGACACCTTATGCTCTGAAATCATGAGCAGGCTGATTCATTTGGTGAATTTGGGTAGA 1206

Qy 599 AAGCAGTATGTTTGTGCTGACATTAAGATGTAGGTTATAGATAGGTTTGTAGCTTTAAGTGT 658
Db 1207 AAGCAGTATGTTTGTGCTGACATTAAGATGTAGGTTATAGATAGGTTTGTAGCTTTAAGTGT 1266

Qy 659 ATGTTTATATCTTAAATAAGAAATATAACCTTTTAAAGCTATTTCCACCTCTCCCTCCCA 718
Db 1267 ATGTTTATATCTTAAATAAGAAATATAACCTTTTAAAGCTATTTCCACCTCTCCCTCCCA 1326

Qy 719 GCCTATCTCAAACTGGTGGATATATGAGAGATCTTGAAGAAGTAAATAAACCCTTCA 778
Db 1327 GCCTATCTCAAACTGGTGGATATATGAGAGATCTTGAAGAAGTAAATAAACCCTTCA 1386

Qy 779 CTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTATTTAAT 838
Db 1387 CTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTATTTAAT 1446

Qy 839 ACTTACCTCTTATTTCTGAATCAGTTGTGAACCTGTTGCCTATGTTTTCAGAGTTTGA 898
Db 1447 ACTTACCTCTTATTTCTGAATCAGTTGTGAACCTGTTGCCTATGTTTTCAGAGTTTGA 1506

Qy 899 ACCTCMGTGAATTCATTTTAAATCTGCTATTTCTGAGAAGCATTTGAATGAATTTCTAA 958
Db 1507 ACCTCMGTGAATTCATTTTAAATCTGCTATTTCTGAGAAGCATTTGAATGAATTTCTAA 1566

Qy 959 CAAGAAGACTCATCTGTAGCTGTTTGTGACTCTCTATGAGCCCATTAAGGGTCTGTGCT 1018
Db 1567 CAAGAAGACTCATCTGTAGCTGTTTGTGACTCTCTATGAGCCCATTAAGGGTCTGTGCT 1625

Qy 1019 TAGCATTAACAAAATAAGTTTATAGTAAAGCAATGTATTAATTTTTCATGGA 1078
Db 1626 TAGCATTAACAAAATAAGTTTATAGTAAAGCAATGTATTAATTTTTCATGGA 1685

Qy 1079 GGGCTTTAAATTTGCTCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTGAAC 1138
Db 1686 GGGCTTTAAATTTGCTCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTGAAC 1745

Qy 1139 TGCTTTTACGGAGATTAATTAATGTTTATAAATAGTTTGGGGGGAATTAATTTGCAAA 1198
Db 1746 TGCTTTTACGGAGATTAATTAATGTTTATAAATAGTTTGGGGGGAATTAATTTGCAAA 1805

Qy 1199 GAGGATTAATTTATTTACGTCTTCTGTTATTCAGAAATAAGAGAGAACTACCTGCA 1258
Db 1806 GAGGATTAATTTATTTACGTCTTCTGTTATTCAGAAATAAGAGAGAACTACCTGCA 1865

Qy 1259 TATTCAGAGTTGTACCTTAACATTTGGTGAAACATTTTCTAAGATTTTCAAAGGAAT 1318
Db 1866 TATTCAGAGTTGTACCTTAACATTTGGTGAAACATTTTCTAAGATTTTCAAAGGAAT 1925

Qy 1319 ATGTGTAATTTGAGAAATCATACCACTGCTTAACCTGTTGTAACAACTGTTCTTAAAT 1378
Db 1926 ATGTGTAATTTGAGAAATCATACCACTGCTTAACCTGTTGTAACAACTGTTCTTAAAT 1985

Qy 1379 AAGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 1413
Db 1986 AAGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 2020

RESULT 4

BC033650
LOCUS
DEFINITION
Homo sapiens hypothetical protein FLJ14431, mRNA (cDNA clone MGC:44889 IMAGE:5574637), complete cds.
ACCESSION
BC033650
VERSION
BC033650.1
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
REFERENCE
1 (bases 1 to 3452)
AUTHORS
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,I.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butcherfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerker, A., Schein, J.E., Jones, S.J. and Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 3452)
Strausberg, R.
Direct Submission
Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-xemail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgrl.nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 69 Row: n Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 14249445.

FEATURES

source
Location/Qualifiers
1. .3452
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:44889 IMAGE:5574637"
/tissue_type="Duodenum, adenocarcinoma"
/clone_lib="NIH_MGC_88"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1. .3452
/gene="FLJ14431"
/db_xref="LocusID:84869"
159. .872
/codon_start=1
/product="FLJ14431 protein"
/protein_id="AAH33650.1"
/db_xref="GI:21707067"
/db_xref="LocusID:84869"
/translation="MDKVCVFGSGRIGRAVQLMARKGYRLAVIARNLEGAKAAG DLGGDLAFSCDVAKHEVDQNTFEELKHLGRVFNVAAGINRDLVLRKTKEDMVS QLHNLGLSMILTKCAAMRTMIOQGGSI NVNGSIVGLKNGSGSVYSASKGGLVFSR ALAKVARKKIRVNVVAPGVFMTDKLKEHLKKNIPILGRGETIEVAHVFFLE SPYITGHVLVVDGGLQLIL"
159. .857
misc_feature

/notes="adh_short; Region: short chain dehydrogenase. This family contains a wide variety of dehydrogenases"
/db_xref="CDD:pfam00106"
Query Match 95.4%; Score 1368.4; DB 9; Length 3452;
Best Local Similarity 99.3%; Pred. No. 6.5e-217;
Matches 1410; Conservative 5; Mismatches 1; Indels 4; Gaps 4;
ORIGIN
QY 1 CATTAAACTCTTTTATCGGGATAGTATGATATTTTCAATGTCACCTCCATTCATGTGA 60
Db 2034 CATTAAACTCTTTTATCGGGATAGTATGATATTTTCAATGTCACCTCCATTCATGTGA 2093
QY 61 TTTGGAGCTGACAGTTATTTTGTGTGAAGCAGAGATTTAAATTTTATATGAAAGTCAGTGC 120
Db 2094 TTTGGAGCTGACAGTTATTTTGTGTGAAGCAGAGATTTAAATTTTATATGAAAGTCAGTGC 2153
QY 121 AAAATTATGAATAGGATATCTAATAATAACAAAGTAATAACAAAGTCAAAAGTCAGAGTGT 180
Db 2154 AAAATTATGAATAGGATATCTAATAATAACAAAGTAATAACAAAGTCAAAAGTCAGTGT 2213
QY 181 CTAATAAATAAATTCCTGGTTCCTTAAATAATTTTAAATTTATC-TTGAATAAGTTTTC 239
Db 2214 CTAATAAATAAATTCCTGGTTCCTTAAATAATTTTAAATTTATC-TTGAATAAGTTTTC 2273
QY 240 TTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATTA 299
Db 2274 TTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATTA 2333
QY 300 ACAAATTGCTATTAATGCMGA-GTGGTAATATACAGAAATTTATCAGGCATTTACCAAG 358
Db 2334 ACAAATTGCTATTAATGCMGACGTGGTAATATACAGAAATTTATCAGGCATTTACCAAG 2393
QY 359 TCTAGGCACATATAGAAATGACGACTCAGAAATGTTTCAATGTAGTAGTTGATGCTTG 418
Db 2394 TCTAGGCACATATAGAAATGACGACTCAGAAATGTTTCAATGTAGTAGTTGATGCTTG 2453
QY 419 TAAAGTAGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGCTGTSTCAATGCT 478
Db 2454 TAAAGTAGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGCTGTCTCAATGCT 2513
QY 479 GGCTTTGGTACTGCTTCCSCATTATGGCAGGCCATTACGTCTTGAGTTTCTTCT 538
Db 2514 GGCTTTGGTACTGCTTCCSCATTATGGCAGGCCATTACGTCTTGAGTTTCTTCT 2573
QY 539 CTGCACACCTTATGCTCTGAAATCATGAGCGAGCTGATTCAATTTGGTGAATTTGGGTAGA 598
Db 2574 CTGCACACCTTATGCTCTGAAATCATGAGCGAGCTGATTCAATTTGGTGAATTTGGGTAGA 2633
QY 599 AAGCAGTATGTTTGTGACATTAAAGATGTAGGTATAGATAGTTTAGCCTTTAAGTGT 658
Db 2634 AAGCAGTATGTTTGTGACATTAAAGATGTAGGTATAGATAGTTTAGCCTTTAAGTGT 2693
QY 659 ATGTTTTATCTTTTAAATAAGAAATATAACCTTTTAAAGCTTATCCACCTCTCCGCCA 718
Db 2694 ATGTTTTATCTTTTAAATAAGAAATATAACCTTTTAAAGCTTATCCACCTCTCCGCCA 2753
QY 719 GCCTATCTCAAACTGCTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCA 778
Db 2754 GCCTATCTCAAACTGCTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCA 2813
QY 779 CTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTGTAGTAAATTTGTAATTAAT 838
Db 2814 CTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTGTAGTAAATTTGTAATTAAT 2873
QY 839 ACTTACTTCTATTCTTGAATCAGTTGTGAACCTGTTCCTTATGTTTCAGAGTTTGAAGA 898
Db 2874 ACTTACTTCTATTCTTGAATCAGTTGTGAACCTGTTCCTTATGTTTCAGAGTTTGAAGA 2933
QY 899 ACCTCMGTGAATTCATTTTAAATACTGCTATTTCTGAGAAGCATTTGAATGAATTTCTTA 958
Db 2934 ACCTCAGTGAATTCATTTTAAATACTGCTATTTCTGAGAAGCATTTGAATGAATTTCTTA 2993

```

QY 959 CAGAAGACTCATCTGTAGCTGTTTCTGCTGACTGCTCCTATGAGCCCATAGGGTTCCTGTGCT 1018
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 2994 CAGAAGACTCATCTGTAGCTGTTTCTGCTGACTGCTCCTATGAG-CCCATAGGGTTCCTGTGCT 3052
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1019 TAGCATTAACAAAAATAAGGTTTATAGGTAAAGCCCAATGTAATATTTTTCATGCGATGGA 1078
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3053 TAGCATTAACAAAAATAAGGTTTATAGGTAAAGCCCAATGTAATATTTTTCATGCGATGGA 3112
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1079 GGGCTTTAAATTTGCTGCTCTTTTTCATATTTTATTCATATTCATATTCATATTCATATTCAT 1138
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3113 GGGCTTTAAATTTGCTGCTCTTTTTCATATTTTATTCATATTCATATTCATATTCATATTCAT 3172
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1139 TGCTTTTTCAGGAGATAATATATATGTTATATATTTAGTTTTCGGGGGGAATATTCGTGCAA 1198
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3173 TGCTTTTTCAGGAGATAATATATATGTTATATATTTAGTTTTCGGGGGGAATATTCGTGCAA 3232
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1199 GAGGATAATTTAAATTTACGTCCTCTGTTATTCAGATAAAGAGAGAGACTACGCTGCA 1258
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3233 GAGGATAATTTAAATTTACGTCCTCTGTTATTCAGATAAAGAGAGAGACTACGCTGCA 3292
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1259 TATTCAGAGTTGTACCTTAACATTCGTGAACA-TTTTCTTAAGATTTCATAAGGAA 1317
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3293 TATTCAGAGTTGTACCTTAACATTCGTGAACA-TTTTCTTAAGATTTCATAAGGAA 3352
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1318 TATGCTAAATTTGAGAAATCATACCTGCTCCTAACTGTTGTAACAACTGTTCTTTAAA 1377
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3353 TATGCTAAATTTGAGAAATCATACCTGCTCCTAACTGTTGTAACAACTGTTCTTTAAA 3412
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 1378 TAAAGTATTAAATGATTTTAAAAA-AAAAAAAAAAAAA 1417
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 3413 TAAAGTATTAAATGATTTTAAAAA-AAAAAAAAAAAAA 3452
Db ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 5
AC021151/c 175081 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-483A20 from 4, complete sequence.
DEFINITION AC021151
ACCESSION AC021151
VERSION AC021151.8 GI:15145598
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175081)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
93063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 175081)
AUTHORS Kyung, K. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RP11-483A20
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 175081)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 175081)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 175081)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 9, 2001 this sequence version replaced gi:1387722.
----- Genome Center

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0483A20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-36G9. Actual start of this clone is at base position 1 of RP11-483A20; actual end is at base position 175081 of RP11-483A20.

FEATURES	Location/Qualifiers
source	1..175081
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-483A20"
	/clone_lib="RP11-11"
repeat_region	1..3267
	/rpt_family="L1"
repeat_region	3270..3453
	/rpt_family="Alu"
repeat_region	3434..3464
	/rpt_family="(A)n"
repeat_region	3455..4161
	/rpt_family="L1"
misc_feature	7949..7956
	/note="match to EST AW006057 (NID:g5854835) wz8ld04.x1"
	/10197..10489
repeat_region	10480..10501
	/rpt_family="AT-rich"
repeat_region	10495..10665
	/rpt_family="L1"
repeat_region	10731..11051
	/rpt_family="L1"
repeat_region	11052..11576
	/rpt_family="L1"
repeat_region	11564..11598
	/rpt_family="(A)n"
repeat_region	11642..12324

	/rpt_family="Li"	repeat_region	11769..11794	/rpt_family="MaLR"	repeat_region	27440..27490
	/rpt_family="AT_rich"	repeat_region	11794..11819	/rpt_family="L1"	repeat_region	27491..28717
	/rpt_family="AT_rich"	repeat_region	11819..11939	/rpt_family="(T)n"	misc_feature	28718..28742
	/rpt_family="Li"	repeat_region	11939..12335	/notes="match to EST AA954400 (NID:g3118095) oo01f01.s1"	misc_feature	28735..28771
	/rpt_family="Li"	repeat_region	12335..12392	/note="match to EST AW006057 (NID:g5954835) w281d04.xl"	repeat_region	28747..28826
	/rpt_family="Li"	repeat_region	12392..12621	/rpt_family="L1"	misc_feature	29022..29308
	/rpt_family="Alu"	repeat_region	12621..12788	/notes="match to EST BF680971 (NID:g11954866)"	misc_feature	29384..30105
	/rpt_family="Alu"	repeat_region	12788..13096	/note="match to EST BG429938 (NID:g13336444)"		
	/rpt_family="Alu"	repeat_region	13096..13069			
	/rpt_family="(CAAAA)n"	repeat_region	13069..13325			
	/rpt_family="Alu"	repeat_region	13325..13667			
	/rpt_family="Alu"	repeat_region	13667..13444			
	/rpt_family="(TAAA)n"	repeat_region	13444..13484			
	/rpt_family="Alu"	repeat_region	13484..13640			
	/rpt_family="(GAAAA)n"	repeat_region	13640..14249			
	/rpt_family="MER1_type"	repeat_region	14249..14980			
	/rpt_family="Li"	repeat_region	14980..15182			
	/rpt_family="Li"	repeat_region	15182..15260			
	/rpt_family="Li"	repeat_region	15260..15483			
	/rpt_family="MER1_type"	repeat_region	15483..15985			
	/rpt_family="Li"	repeat_region	15985..16105			
	/rpt_family="Li"	repeat_region	16105..16117			
	/rpt_family="Li"	repeat_region	16117..16251			
	/rpt_family="Li"	repeat_region	16251..16288			
	/rpt_family="Li"	repeat_region	16288..16373			
	/rpt_family="Li"	repeat_region	16373..16545			
	/rpt_family="Li"	repeat_region	16545..16550			
	/rpt_family="Li"	repeat_region	16550..18770			
	/rpt_family="Li"	repeat_region	18770..19228			
	/rpt_family="T-rich"	repeat_region	19228..20219			
	/rpt_family="Alu"	repeat_region	20219..20514			
	/rpt_family="Alu"	repeat_region	20514..20515			
	/rpt_family="Li"	repeat_region	20515..22484			
	/rpt_family="Alu"	repeat_region	22484..22584			
	/rpt_family="Alu"	repeat_region	22584..22748			
	/rpt_family="AT_rich"	repeat_region	22748..22883			
	/rpt_family="Li"	repeat_region	22883..23461			
	/rpt_family="MaLR"	repeat_region	23461..23747			
	/rpt_family="Li"	repeat_region	23747..24456			
	/note="match to EST AW902102 (NID:g8066307)"	misc_feature	24456..24488			
	/note="match to EST AW902102 (NID:g8066307)"	misc_feature	24488..24576			
	/rpt_family="Mariner"	repeat_region	24576..25484			
	/rpt_family="Alu"	repeat_region	25484..25758			
	/rpt_family="(A)n"	repeat_region	25758..25796			
	/rpt_family="MIR"	repeat_region	25796..26657			
	/rpt_family="Li"	repeat_region	26657..26890			
	/rpt_family="MaLR"	repeat_region	26890..27206			
	/rpt_family="Alu"	repeat_region	27206..27439			


```
* 9518 1617: gap of unknown length
* 9618 12107: contig of 2490 bp in length
* 12108 12207: gap of unknown length
* 12208 15014: contig of 2807 bp in length
* 15015 15114: gap of unknown length
* 15115 17831: contig of 2717 bp in length
* 17832 17931: gap of unknown length
* 17932 20480: contig of 2549 bp in length
* 20481 20580: gap of unknown length
* 20581 22700: contig of 2120 bp in length
* 22701 22800: gap of unknown length
* 22801 24909: contig of 2109 bp in length
* 24910 25009: gap of unknown length
* 25010 27313: contig of 2304 bp in length
* 27314 27413: gap of unknown length
* 27414 29798: contig of 2385 bp in length
* 29799 29898: gap of unknown length
* 29899 33466: contig of 3568 bp in length
* 33467 33566: gap of unknown length
* 33567 36679: contig of 3113 bp in length
* 36680 36779: gap of unknown length
* 36780 40472: contig of 3693 bp in length
* 40473 40572: gap of unknown length
* 40573 42683: contig of 2111 bp in length
* 42684 42783: gap of unknown length
* 42784 46319: contig of 3536 bp in length
* 46320 46419: gap of unknown length
* 46420 49507: contig of 3088 bp in length
* 49508 49607: gap of unknown length
* 49608 53212: contig of 3605 bp in length
* 53213 53312: gap of unknown length
* 53313 57267: contig of 3955 bp in length
* 57268 57367: gap of unknown length
* 57368 60444: contig of 3077 bp in length
* 60445 60544: gap of unknown length
* 60545 65859: contig of 5315 bp in length
* 65860 65939: gap of unknown length
* 65940 69038: contig of 3079 bp in length
* 69039 69138: gap of unknown length
* 69139 73445: contig of 4307 bp in length
* 73446 73545: gap of unknown length
* 73546 78107: contig of 4562 bp in length
* 78108 78207: gap of unknown length
* 78209 82504: contig of 4297 bp in length
* 82505 82604: gap of unknown length
* 82605 86976: contig of 4372 bp in length
* 86977 87076: gap of unknown length
* 87077 91315: contig of 4239 bp in length
* 91316 91415: gap of unknown length
* 91416 95084: contig of 3669 bp in length
* 95085 95184: gap of unknown length
* 95185 102348: contig of 7164 bp in length
* 102349 102448: gap of unknown length
* 102449 105973: contig of 3525 bp in length
* 105974 106073: gap of unknown length
* 106074 112891: contig of 6818 bp in length
* 112892 112991: gap of unknown length
* 112992 116221: contig of 3230 bp in length
* 116222 116321: gap of unknown length
* 116322 119798: contig of 3477 bp in length
* 119799 119898: gap of unknown length
* 119899 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140738: gap of unknown length
* 140739 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155057: contig of 9416 bp in length
* 155058 155157: gap of unknown length

* 155158 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174892: contig of 10435 bp in length
* 174893 174992: gap of unknown length
* 174993 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 216536: gap of unknown length
* 216537 232917: contig of 16380 bp in length
* 232918 233016: gap of unknown length
* 233017 255437: contig of 22421 bp in length
* 255438 255537: gap of unknown length
* 255538 281662: contig of 26125 bp in length.

FEATURES             Location/Qualifiers
     source            1..281662
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="12"
                        /clone="RP11-19D19"

ORIGIN
Query Match      41.2%; Score 591; DB 2; Length 281662;
Best Local Similarity 98.1%; Pred. No. 9.8e-89;
Matches 617; Conservative 2; Mismatches 7; Indels 3; Gaps 2;

QY 773 CCTTCACTGCTCCACCTCCAGGTGAATCCCGCCCACTCCCACTGACCTAGTAGAATTGTAA 832
DB 33860 CCTCACCGCTCCCTCCACGTGAATCCCGCCCACTCCCACTGACCTAGTAGAATTGTAA 33919

QY 833 TTATAATAC--TTACCTCTTATTTCTGAATCAGTGTGAACTGTGCTTTATGTTTCAGAR 890
DB 33920 TTATAACGTAGCTTCTATTTCTGAATCAGTGTGAACTGTGCTTTATGTTTCAGAR 33979

QY 891 GTTTAAGAACCTCMGTGGAATTCATTTTTTAAATCTGCTATTCTGAGAAGCAATTTGAATGA 950
DB 33980 GTTTAAGAACCTCAGTGAATTCATTTTTTAAATCTGCTATTCTGAGAAGCAATTTGAATGA 34039

QY 951 ATTCTTAAACAGAGACTCATCTGTAGCTGTTTGTGCTGATCCCTATGAGCCCAATAGGGT 1010
DB 34040 ATTCTTAAACAGAGACTCATCTGTAGCTGTTTGTGCTGATCCCTATGAG--CCCAATAGGGT 34098

QY 1011 TCTGTGCTTAGCAATTAAACAAATAAGGTTTATAGGTAAGCCCAATCTATTAAATTTTTTT 1070
DB 34099 TCTGTGCTTAGCAATTAAACAAATAAGGTTTATAGGTAAGCCCAATCTATTAAATTTTTTT 34158

QY 1071 TGCATCGAGGGCTTTAAAATTTGTCCTCTTTTTCATATTTTATTCATATTTCAATTTATGG 1130
DB 34159 TGCATCGAGGGCTTTAAAATTTGTCCTCTTTTTCATATTTTATTCATATTTCAATTTATGG 34218

QY 1131 TTGTGAACCTGCTTTTATAGGGAGTAAATATATGTTTAAATTTAGTTTGGGGGGAATAAT 1190
DB 34219 TTGTGAACCTGCTTTTATAGGGAGTAAATATATGTTTAAATTTAGTTTGGGGGGAATAAT 34278

QY 1191 TGTGCAAGAGGATAATTTAAATTTACGTGCTCTCTGTTTATTCAGATAAAGAGAGAGAGACT 1250
DB 34279 TGTGCAAGAGGATAATTTAAATTTACGTGCTCTCTGTTTATTCAGATAAAGAGAGAGAGACT 34338

QY 1251 AGCTGCATATTCAGAGATTGTACCTTAACATTGGTGAACATTTTTTTCTAAGATTTTCA 1310
DB 34339 AGCTGCATATTCAGAGATTGTACCTTAACATTGGTGAACATTTTTTTCTAAGATTTTCA 34398

QY 1311 AAAGGAATATGCTAAATTTGAGAAATCATAAACCACTGCTCTTAACCTTGGTAAACAACTGT 1370
DB 34399 AAAGGAATATGCTAAATTTGAGAAATCATAAACCACTGCTCTTAACCTTGGTAAACAACTGT 34458

QY 1371 TCTTAATAAAGTATTTAATGATTTTAA 1399
DB 34459 TCTTAATAAAGTATTTAATGATTTTAA 34487
```

```

RESULT 7
AC117374/c
LOCUS      AC117374      281662 bp      DNA      linear      HTG 02-MAY-2002
DEFINITION Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
PROGRESS ***, 48 unordered pieces.
ACCESSION  AC117374 AC011694
VERSION     AC117374.1 GI:20127352
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 281662)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Ausbrooks,S.H., Amaratunge,H.C., Are,J.R., Ayie,M., Banks,T.,
            Barbara,J., Benton,J., Bmaga,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Neal,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,N.,
            Moser,M., Nial,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokkwo,S., Ogih,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Piekens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
            Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Ugmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstein,G. and Gibbs,R.
            Direct Submission
            2 (bases 1 to 281662)
            Worley,K.C.
            Direct Submission
            Submitted (10-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 281662)
            Worley,K.C.
            Direct Submission
            Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Apr 28, 2002 this sequence version replaced gi:10045385.
            -----
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -----
            Project Information
            Center project name: HCND
            Center clone name: RP11-19D19
            -----
            Summary Statistics
            Sequencing vector: Plasmid;
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 273063 bases at least Q40
            Consensus quality: 294336 bases at least Q30
            Consensus quality: 306942 bases at least Q20
            -----
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 48 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            1 2464: contig of 2464 bp in length
            * 2465: gap of unknown length
            * 2565: contig of 2039 bp in length
            * 4604: gap of unknown length
            * 4704: contig of 2410 bp in length
            * 7114: gap of unknown length
            * 7214: contig of 2304 bp in length
            * 9518: gap of unknown length
            * 9617: contig of 2490 bp in length
            * 12108: gap of unknown length
            * 12207: contig of 2807 bp in length
            * 15015: gap of unknown length
            * 15115: contig of 2717 bp in length
            * 17832: gap of unknown length
            * 17932: contig of 2549 bp in length
            * 20480: gap of unknown length
            * 20581: contig of 2120 bp in length
            * 22700: gap of unknown length
            * 22801: contig of 2109 bp in length
            * 24910: gap of unknown length
            * 25010: contig of 2304 bp in length
            * 27113: gap of unknown length
            * 27414: contig of 2385 bp in length
            * 29798: gap of unknown length
            * 29898: gap of unknown length
            * 33466: contig of 3568 bp in length
            * 33566: gap of unknown length
            * 36679: contig of 3113 bp in length
            * 36779: gap of unknown length
            * 40472: contig of 3693 bp in length
            * 40572: gap of unknown length
            * 42683: contig of 2111 bp in length
            * 42783: gap of unknown length
            * 46319: contig of 3536 bp in length
            * 46419: gap of unknown length
            * 49507: contig of 3088 bp in length
            * 49607: gap of unknown length
            * 53212: contig of 3605 bp in length
            * 53312: gap of unknown length
            * 57267: contig of 3955 bp in length
            * 57367: gap of unknown length
            * 60444: contig of 3077 bp in length
            * 60544: gap of unknown length
            * 65859: contig of 5315 bp in length
            * 65953: gap of unknown length
            * 69038: contig of 3079 bp in length
            * 69138: gap of unknown length
            * 69339: contig of 4307 bp in length
            * 73446: gap of unknown length
            * 73546: contig of 4562 bp in length
            * 78107: gap of unknown length
            * 78207: gap of unknown length
            * 82504: contig of 4297 bp in length
            * 82604: gap of unknown length

```

```
* 82605 86976: contig of 4372 bp in length
* 86977 87076: gap of unknown length
* 87077 91315: contig of 4239 bp in length
* 91316 91415: gap of unknown length
* 91416 95084: contig of 3669 bp in length
* 95085 95184: gap of unknown length
* 95185 102348: contig of 7164 bp in length
* 102349 102448: gap of unknown length
* 102449 105973: contig of 3525 bp in length
* 105974 106073: gap of unknown length
* 106074 112891: contig of 6818 bp in length
* 112892 112991: gap of unknown length
* 112992 116221: contig of 3230 bp in length
* 116222 116321: gap of unknown length
* 116322 119798: contig of 3477 bp in length
* 119799 119898: gap of unknown length
* 119899 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140737: gap of unknown length
* 140738 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155057: contig of 9416 bp in length
* 155058 155157: gap of unknown length
* 155158 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174892: contig of 10435 bp in length
* 174893 174992: gap of unknown length
* 174993 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 216536: gap of unknown length
* 216537 232916: contig of 16380 bp in length
* 232917 233016: gap of unknown length
* 233017 255437: contig of 22421 bp in length
* 255438 255537: gap of unknown length
* 255538 281662: contig of 26125 bp in length.
FEATURES
    source
        1..281662
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="12"
            /clone="RP11-19D19"
ORIGIN
    Query Match      38.7%; Score 555; DB 2; Length 281662;
    Best Local Similarity 98.6%; Pred. No. 8.8e-83;
    Matches 577; Conservative 3; Mismatches 3; Indels 2; Gaps 2;
QY 1 CATTAACTCTTTTTCGGGAATAGTATGATATTTTCAATGTCACCTCCATCATCTTGA 60
Db 146226 CATTAACTCTTTTTCGGGAATAGTATGATATTTTTCAGTGTCACTCCATTCATCTTGA 146167
QY 61 TTGGAGCTGACAGTATTTCGTGAAGCAGAGATTATTTTATATTGAAAGTCAGTGC 120
Db 146166 TTGGAGCTGACAGTATTTCGTGAAGCAGAGATTATTTTATATTGAAAGTCAGTGC 146107
QY 121 AAAAAATGAAATAGGATATCTAAATAAATACAAAGTAATAACAAAAGTCAAAGCAGTGT 180
Db 146106 AAAAAATGAAATAGGATATCTAAATAAATACAAAGTAATAACAAAAGTCAAAGCAGTGT 146047
QY 181 CTAATAAAAAATTCGGTTCCTTAAAAATTAATTTAAATTTATC-TTGAAATAGTTTTTC 239
Db 146046 CTAATAAAAAATTCGGTTCCTTAAAAATTAATTTAAATTTATCTTTGAAATAGTTTTTC 145987
QY 240 TTAGATTAACTCAGAGATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 299
Db 145986 TTAGATTAACTCAGAGATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 145927
QY 300 ACAAAATTGCTATTAAATGCAMGA-GTGGTAATATACAGAAATTTATCAGGCATTACCAAG 358
Db 145926 ACAAAATTGCTATTAAATGCAMGA-GTGGTAATATACAGAAATTTATCAGGCATTACCAAG 145867
QY 359 TCTAGGCACATATAGGAATGCAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTG 418
Db 145866 TCTAGGCACATATAGGAATGCAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTG 145807
QY 419 TAAGGTAGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGCTGTSTCAATTTGCT 478
Db 145806 TAAGGTAGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGCTGTCTCAATTTGCT 145747
QY 479 GGCTTTTGCTACCTGTACTTCCSATTATGCAGGCCAATTCAGTCTTCAGTTTCTTCT 538
Db 145746 GGCTTTTGCTACCTGTACTTCCSATTATGCAGGCCAATTCAGTCTTCAGTTTCTTCT 145687
QY 539 CTGGACACCTTATGCTCTGAAATCATGACGAGGCTGATTCAATT 583
Db 145686 CTGGACACCATATGCTCTGAAATCATGACGAGGCTGAATCAATT 145642
RESULT 8
AX408738/c
LOCUS AX408738 448 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1385 from Patent WO0229103.
ACCESSION AX408738
VERSION AX408738.1 GI:21441443
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 1385 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
    Location/Qualifiers
        1..448
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="EMBL/GenBank Accession No. AA491000"
ORIGIN
    Query Match      29.7%; Score 425.8; DB 6; Length 448;
    Best Local Similarity 98.2%; Pred. No. 7.5e-61;
    Matches 441; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 965 GACTCATCTGTAGCTGTTTGTGACTCTCTATGAGCCCATAGGGTCTGTGCTTAGCAT 1024
Db 448 GACTCATCTGTAGCTGTTTGTGACTCTCTATGAGCCCATAGGGTCTGTGCTTAGCAT 390
QY 1025 TAACAAAATAAGGTTTATAGGTAAGCCCAATGTATTAATTTTTTTTGCATGGAGGCTT 1084
Db 389 TAACAAAATAAGGTTTATAGGTAAGCCCAATGTATTAATTTTTTTTGCATGGAGGCTT 330
QY 1085 TAAATTTGCTCTTTTTCATATTATTATTCATATTCAATTTTATGTTTGTAACTGCTTT 1144
Db 329 TAAATTTGCTCTTTTTCATATTATTATTCATATTCAATTTTATGTTTGTAACTGCTTT 270
QY 1145 TTAGGAGATAATATATGTTTATTAATAGTTTGGGGGAATAATTTGTCAAAAGGAT 1204
Db 269 TTAGGAGATAATATATGTTTATTAATAGTTTGGGGGAATAATTTGTCAAAAGGAT 210
QY 1205 AATTTAAATTTAGTGTCTTCTGTTATTTCAGATAAAGAGAGAGACTACGCTGCATATTC 1264
Db 209 AATTTAAATTTAGTGTCTTCTGTTATTTCAGATAAAGAGAGAGACTACGCTGCATATTC 150
QY 1265 AGAGTTGTACCTTAACATTGGTGAACAATTTTTCTAAGATTTTCAAAGGAATATGTGT 1324
```



```

ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 184359)
AUTHORS
Harris,A., Meyer,R. and Spalding,L.
TITLE
The sequence of Mus musculus BAC clone RP24-134L8
REFERENCE
2 (bases 1 to 184359)
AUTHORS
Wilson,R.
TITLE
Sequencing of Mus musculus
JOURNAL
Unpublished (2001)
REFERENCE
3 (bases 1 to 184359)
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
4 (bases 1 to 184359)
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
5 (bases 1 to 184359)
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
Submitted (02-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
6 (bases 1 to 184359)
AUTHORS
Wilson,R.
TITLE
Direct Submission
JOURNAL
Submitted (11-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 2, 2002 this sequence version replaced gi:22475821.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
-----
Center project name: M_BB0134L08
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC117196.
Location/Qualifiers
1. .184359
/organism="Mus musculus"
/mol_type="genomic DNA"

FEATURES
source
/db_xref="taxon:10090"
/chromosome="8"
/map="8"
/clone="RP24-134L8"
/clone_lib="RPCI-24"
/repeat_region
289..527
/rpt_family="B4"
1994..2232
/rpt_family="ERVK"
2383..3887
/rpt_family="L1"
3888..4101
/rpt_family="ERVK"
4327..4514
/rpt_family="MaLR"
4652..4817
/rpt_family="RMER15"
5330..5642
/rpt_family="MaLR"
6281..6619
/rpt_family="RMER10B"
6759..7020
/rpt_family="B4"
7183..7363
/rpt_family="B2"
7436..7681
/rpt_family="MaLR"
7701..7824
/rpt_family="MaLR"
9897..10028
/rpt_family="Alu"
10171..10305
/rpt_family="Alu"
10314..10449
/rpt_family="B4"
11501..11617
/rpt_family="Alu"
12345..12438
/rpt_family="BCI_MM"
12471..12629
/rpt_family="B4"
12684..12778
/rpt_family="B4"
12914..13275
/rpt_family="L1"
13283..14578
/rpt_family="L1"
14627..15630
/rpt_family="L1"
15739..15857
/rpt_family="ERV1"
16247..16375
/rpt_family="Alu"
16501..16615
/rpt_family="Alu"
16655..16713
/rpt_family="Alu"
17787..17878
/rpt_family="Alu"
17950..18029
/rpt_family="Alu"
18461..18609
/rpt_family="B4"
18816..18908
/rpt_family="Alu"
19317..19666
/rpt_family="MaLR"
19739..19928
/rpt_family="B2"
20490..20679
/rpt_family="B2"
21147..21374
/rpt_family="L1"

```


of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 210911)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23816560.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GSMW
 Center clone name: CH230-145I15

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 185401 bases at least Q40

Consensus quality: 188841 bases at least Q30

Consensus quality: 191364 bases at least Q20

Estimated insert size: 191507; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 210911: contig of 210911 bp in length.

FEATURES

source
 1. .210911
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-145I15"
 misc_feature
 1. .1076
 /note="wgs contig"
 misc_feature
 complement(209249..210107)
 /note="clone boundary"
 clone_end:T7
 site:
 end_sequence:RWBAAS6TUC"

ORIGIN

Query Match 5.7%; Score 82; DB 2; Length 210911;
 Best Local Similarity 59.5%; Pred. No. 0.00016;
 Matches 219; Conservative 0; Mismatches 130; Indels 19; Gaps 4;
 QY 943 TTGAATGAATCTTAAACAAGAGACTCACTGTAGCTGTTGCTGCTATGAGCCCC 1002
 Db TTTAGCCAATCTTATACAGAAGACTTATTATGCTGCCACCTCTCTACA---GCTC 72335
 QY 1003 ATAAAGGTTCTGCTAGCATTAACAAATAGGTTTATAGTAAAGCCCAATGTATTAA 1062

Db 72334 ATGGTCCAGCATGGTTAGCAATTTGAAGGAAAAGGTTGGGGCTGAAGCCAGAGTACTCA 72275
 QY 1063 TTTTCTTTTTCATGGAGGGCTTTAAATTTGCTCTCTTTTTCATATTTTATTCATATTC 1122
 Db 72274 GTTTT-----ACGCTGTTTTTAAATATATGCTCTCTTTTCATATTTTATTCATATTC 72220
 QY 1123 ATTAT---GGTTTGAACCTGCTTTTGGGAGATAATATATGTTATAAAATAGTTTTC 1179
 Db 72219 ATTATGAGGGGCTGTGTCAGCTTTATACAGATATGATTATATGTTGCAATTA----- 72166
 QY 1180 GGGGGAATATTTGCAAGAGAGATATTTAATTTAGTCTCTCTGTTATTCAGATAAA 1239
 Db 72165 -GGGGATAATTTATGCAAGAGAGCTGAGCTGATTATACCAATAGTGA 72107
 QY 1240 GAGAGAAGACTACGCTGCATTTTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTTC 1299
 Db 72106 GAGAGACAGATTTTGGGTATTTCAAGAGTTGAGACTAGTATTTATGCGGAATGTTG 72047
 QY 1300 TAAGATTT 1307
 Db 72046 AGATTTT 72039

RESULT 13

AC009412

LOCUS

DEFINITION

AC009412

AC009412.6

GI:14190769

HTG.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 212827)

Sulston,J.E. and Waterston,R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

9847074

2 (bases 1 to 212827)

Hou,S., Maupin,R., Haakenson,W., Gregory,S. and Belter,E.

The sequence of Homo sapiens BAC clone RP11-425P5

Unpublished

3 (bases 1 to 212827)

Waterston,R.H.

Direct Submission

Submitted (21-AUG-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 212827)

Waterston,R.H.

Direct Submission

Submitted (23-MAY-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 212827)

Waterston,R.

Direct Submission

Submitted (07-NOV-2001) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On May 23, 2001 this sequence version replaced gi:13431187.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: sapiens@wustl.edu

----- Summary Statistics

Center project name: H_NH0425P05

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osceogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanesi, J.J., and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTD-2195F2, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

FEATURES

repeat_region	8720..9031	/rpt_family="Alu"	
misc_feature	9083..9560	/note="similar to EST AA633603 (NID:g2556817) ae66b02.s1"	
repeat_region	9598..9899	/rpt_family="Alu"	
misc_feature	9871..10098	/note="similar to EST AA699398 (NID:g2702592) zi40a03.s1"	
repeat_region	10746..11044	/rpt_family="L1"	
repeat_region	11075..11101	/rpt_family="AT-rich"	
repeat_region	11102..11372	/rpt_family="Alu"	
repeat_region	11411..11465	/rpt_family="GA-rich"	
repeat_region	11552..11644	/rpt_family="L2"	
repeat_region	11685..12193	/rpt_family="L1"	
repeat_region	12194..12486	/rpt_family="Alu"	
repeat_region	12487..13003	/rpt_family="L1"	
repeat_region	13004..13305	/rpt_family="Alu"	
repeat_region	13306..13501	/rpt_family="L1"	
repeat_region	13503..13892	/rpt_family="L1"	
misc_feature	13893..14224	/note="similar to EST BE061102 (NID:g8405752)"	
repeat_region	14244..14544	/rpt_family="Alu"	
repeat_region	14546..14828	/rpt_family="Alu"	
repeat_region	15091..15189	/rpt_family="MIR"	
repeat_region	15672..16449	/rpt_family="Achoho"	
repeat_region	16640..16844	/rpt_family="Alu"	
repeat_region	16917..16959	/rpt_family="Achoho"	
repeat_region	16960..17287	/rpt_family="Alu"	
repeat_region	17288..17572	/rpt_family="Achoho"	
misc_feature	17610..17938	/note="similar to EST BF996427 (NID:g12402750)"	
repeat_region	17793..18068	/rpt_family="Alu"	
repeat_region	18251..18537	/rpt_family="Alu"	
misc_feature	18485..18884	/note="similar to EST BE062347 (NID:g8406997)"	
misc_feature	18661..18809	/note="similar to EST BG186544 (NID:g13708231)"	
repeat_region	18957..19160	/rpt_family="L1"	
repeat_region	19314..19528	/rpt_family="L1"	
repeat_region	19819..20135	/rpt_family="Alu"	
repeat_region	21101..21269	/rpt_family="L1"	
repeat_region	21882..22203	/rpt_family="L1"	
misc_feature	22429..22797	/note="similar to EST BF993478 (NID:g12399801)"	
repeat_region	22717..23007	/rpt_family="L1"	
misc_feature	22993..23191		

```

repeat_region
  rpt_family="Alu"
misc_feature
  note="similar to EST BF767044 (NID:gl2115035)"
repeat_region
  rpt_family="Alu"
misc_feature
  note="similar to EST AW105611 (NID:g5076346) xd49g02.x1"
repeat_region
  rpt_family="Alu"
misc_feature
  note="similar to EST 25677"

```

Query Match	5.4%;	Score 77.4;	DB 9;	Length 212827;
Best Local Similarity	64.5%;	Pred. No. 0.00091;		
Matches 180;	Conservative 0;	Mismatches 91;	Indels 8;	Gaps 4;
QY	1136	AACTGCTTTT	TAGGAGATAA	TATATATGTTATATAAATTAGT---TTTGGGGGAATAAATG 1192
Db	79556	ACCTGCTCTT	TAAAGATAGAA	TGTGTCTTAAATCATTTGGCCATTGTGAATAATATTTA 79615
QY	1193	TGCAAGAGAGAT	TAATTAAATTTAC	GTGCTTCTGTTATTCAGAAATAAGAGAGAGAACTAC 1252
Db	79616	TGCAAGAGAGATT	TATGTAATTTATC	AGCTTCTATTATCCAGAAATGAAG---GAAGAAATAT 79672
QY	1253	GCTGCATATT	TCAGAGTGTGTAC	CTTAAACATTTGGTGAACAATTTTTTCTAAGATTTTTCAA 1312
Db	79673	TCTAGATATT	TCAGGACCAATGC	CTT-ACATGGCTGAATATTATATCCCAACTTTTTTATT 79731
QY	1313	AGCAATATGTA	TAAATTTGAGTAAT	CATATACCACTGCTCTAACTTGTGTAAACAACTGTTC 1372
Db	79732	GTAATATGTGTA	AACTGGAAATCAT-AC	TACTGTGCTAAACGTTAAAAACGTTTTTG 79790
QY	1373	TTAAATAAAGTA	TTTTTAATGATTTT	TAAAAAATAAAAAA 1411
Db	79791	TTTTGTTTTTGT	TTTTTTTTTTTTTAA	AGACCAAGATGTAAA 79829

RESULT 14	
AC099821/c	
LOCUS	72832 bp DNA linear HTG 22-NOV-2001
DEFINITION	Homo sapiens chromosome 15 clone RP11-315A19 map 15, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC099821
VERSION	AC099821.1 GI:17047157
KEYWORDS	HTG; HTGS PHASE0.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 72832) Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens chromosome 15, clone RP11-315A19
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 72832) Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Felreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamarez,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE	Direct Submission

JOURNAL
COMMENT

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L12444
Center clone name: 315 A 19

* NOTE: This record contains 94 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

*	1	745:	contig of 745 bp in length
*		845:	gap of 100 bp
*	746	846:	contig of 782 bp in length
*	846	1627:	gap of 100 bp
*	1728	1727:	gap of 100 bp
*	1628	2526:	contig of 799 bp in length
*	2527	2626:	gap of 100 bp
*	2627	3387:	contig of 761 bp in length
*	3388	3487:	gap of 100 bp
*	3488	4284:	contig of 797 bp in length
*	4285	4384:	gap of 100 bp
*	4385	5102:	contig of 718 bp in length
*	5103	5202:	gap of 100 bp
*	5203	5952:	contig of 750 bp in length
*	5953	6052:	gap of 100 bp
*	6053	6866:	contig of 814 bp in length
*	6867	6966:	gap of 100 bp
*	6967	7732:	contig of 766 bp in length
*	7733	7832:	gap of 100 bp
*	7833	8578:	contig of 746 bp in length
*	8579	8678:	gap of 100 bp
*	8679	9465:	contig of 787 bp in length
*	9466	9565:	gap of 100 bp
*	9566	10262:	contig of 697 bp in length
*	10263	10362:	gap of 100 bp
*	10363	11133:	contig of 771 bp in length
*	11134	11233:	gap of 100 bp
*	11234	11982:	contig of 749 bp in length
*	11983	12082:	gap of 100 bp
*	12083	12832:	contig of 750 bp in length
*	12833	12932:	gap of 100 bp
*	12933	13697:	contig of 765 bp in length
*	13698	13797:	gap of 100 bp
*	13798	14553:	contig of 756 bp in length
*	14554	14653:	gap of 100 bp
*	14654	15423:	contig of 770 bp in length
*	15424	15523:	gap of 100 bp
*	15524	16287:	contig of 764 bp in length
*	16288	16387:	gap of 100 bp
*	16388	17135:	contig of 748 bp in length
*	17136	17235:	gap of 100 bp
*	17236	18001:	contig of 766 bp in length
*	18002	18101:	gap of 100 bp
*	18102	18876:	contig of 775 bp in length
*	18877	18976:	gap of 100 bp
*	18978	19558:	contig of 782 bp in length
*	19559	20621:	contig of 763 bp in length
*	19859	20622:	contig of 763 bp in length
*	20622	20721:	gap of 100 bp

```
* 20722 21529: contig of 808 bp in length
* 21530 21629: gap of 100 bp
* 21630 22393: contig of 764 bp in length
* 22394 22493: gap of 100 bp
* 22494 23260: contig of 767 bp in length
* 23261 23360: gap of 100 bp
* 23361 24148: contig of 788 bp in length
* 24149 24248: gap of 100 bp
* 24249 25036: contig of 788 bp in length
* 25037 25136: gap of 100 bp
* 25137 25906: contig of 770 bp in length
* 25907 26006: gap of 100 bp
* 26007 26773: contig of 767 bp in length
* 26774 26873: gap of 100 bp
* 26874 27630: contig of 757 bp in length
* 27631 27730: gap of 100 bp
* 27731 28554: contig of 824 bp in length
* 28555 29483: contig of 829 bp in length
* 29484 29583: gap of 100 bp
* 29584 30372: contig of 789 bp in length
* 30373 30472: gap of 100 bp
* 30473 31230: contig of 758 bp in length
* 31231 31330: gap of 100 bp
* 31331 32132: contig of 802 bp in length
* 32133 32232: gap of 100 bp
* 32233 33005: contig of 773 bp in length
* 33006 33105: gap of 100 bp
* 33106 33862: contig of 757 bp in length
* 33863 33962: gap of 100 bp
* 33963 34729: contig of 767 bp in length
* 34730 34829: gap of 100 bp
* 34830 35597: contig of 768 bp in length
* 35598 35697: gap of 100 bp
* 35698 36479: contig of 782 bp in length
* 36480 36579: gap of 100 bp
* 36580 37350: contig of 771 bp in length
* 37351 37450: gap of 100 bp
* 37451 38223: contig of 773 bp in length
* 38224 38977: gap of 100 bp
* 38978 39077: contig of 654 bp in length
* 39078 39863: contig of 786 bp in length
* 39864 39963: gap of 100 bp
* 39964 40757: contig of 794 bp in length
* 40758 40857: gap of 100 bp
* 40858 41620: contig of 763 bp in length
* 41621 41720: gap of 100 bp
* 41721 42451: contig of 731 bp in length
* 42452 42551: gap of 100 bp
* 42552 43301: contig of 750 bp in length
* 43302 43401: gap of 100 bp
* 43402 44183: contig of 782 bp in length
* 44184 44283: gap of 100 bp
* 44284 45037: contig of 754 bp in length
* 45038 45137: gap of 100 bp
* 45138 45897: contig of 760 bp in length
* 45898 45997: gap of 100 bp
* 45998 46781: contig of 784 bp in length
* 46782 46881: gap of 100 bp
* 46882 47677: contig of 796 bp in length
* 47678 47777: gap of 100 bp
* 47778 48535: contig of 758 bp in length
* 48536 48635: gap of 100 bp
* 48636 49385: contig of 750 bp in length
* 49386 49485: gap of 100 bp
* 49486 50288: contig of 803 bp in length
* 50289 50388: gap of 100 bp
* 50389 51159: contig of 771 bp in length
* 51160 51259: gap of 100 bp
* 51260 51995: contig of 736 bp in length
* 51996 52095: gap of 100 bp
* 52096 52865: contig of 770 bp in length
```

```
* 52866 52965: gap of 100 bp
* 52966 53738: contig of 773 bp in length
* 53739 53838: gap of 100 bp
* 53839 54621: contig of 783 bp in length
* 54622 54721: gap of 100 bp
* 54722 55479: contig of 758 bp in length
* 55480 55579: gap of 100 bp
* 55580 56361: contig of 782 bp in length
* 56362 56461: gap of 100 bp
* 56462 57212: contig of 751 bp in length
* 57213 57312: gap of 100 bp
* 57313 58070: contig of 758 bp in length
* 58071 58170: gap of 100 bp
* 58171 58945: contig of 775 bp in length
* 58946 59045: gap of 100 bp
* 59046 59771: contig of 726 bp in length

Query Match 5.3%; Score 76.6; DB 2; Length 72832;
Best Local Similarity 36.3%; Pred. No. 0.0015;
Matches 154; Conservative 0; Mismatches 268; Indels 2; Gaps 1;

Qy 1010 TTCTGCTTAGCATTAAACAAATAAGGTTTATAGGTAAAGCCCAATGTTATTTT 1069
Db 42355 TTTTNNNNNTTTTATTATATAAAAAATAAATANTTAATTTATTTNATTAANNNNNN 42296

Qy 1070 TTGCATGGAGGCTTTAAAAATTTGCTCTCTTTTCATATTTTTCATATTTTATG 1129
Db 42295 NNNNNNNNTTTTNTTANTTTTNTTNNNNNTTNNNNNTTNNNNNTTNNNNNTT 42238

Qy 1130 GTTTGTAACGCTTTTATAGGAGATAATATATGTTATATAATAGTTTGGGGGAATAA 1189
Db 42237 TATNNNTANNNTTTTNTTANNNTTAAANATAATTTTNTTANNNTTNNNNNNNTN 42178

Qy 1190 TTGTGCAAGAGGATAATTTAAATTTACGTCCTCTGTTATTCAGATAAAGAGAGAC 1249
Db 42177 NNNNNNANTNTNANNNNNNNNAATTTTNNNNNTTTTNNNNNTTNNNNNTTNNNN 42118

Qy 1250 TAGCTGCATATCAAGAGTTGTACCTTAACATTTGGTGAACATTTTCTTCTAAGATTTC 1309
Db 42117 NANNNTNATTTTAAATNTNNAANTNNNNNNAATNNNNNNNNNNNNNNNNNNNTT 42058

Qy 1310 AAAAGGAATATGTAAATTTAGAAATCATACCACTGCTCTAACTTGGTAAACAAACTG 1369
Db 42057 AAAANNTTANTTTTTTTTNTANNNANNAANNAANACCNNANATAAAAAANNC 41998

Qy 1370 TTCTTAAATAAGTATTATGATTTTAAAAAATAAAAAAATAAAAAAATAAAAAA 1429
Db 41997 NCCNCAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 41938

Qy 1430 AAAA 1433
Db 41937 TATA 41934
```

RESULT 15

AC116984 1/c

WPCOMMENT

Sequence split into 6 fragments LOCUS AC116984 Accession AC116984

Fragment Name	Begin	End
AC116984.0	1	110000
AC116984.1	100001	210000
AC116984.2	200001	310000
AC116984.3	300001	410000
AC116984.4	400001	510000
AC116984.5	500001	541399

Continuation (2 of 6) of AC116984 from base 100001 (AC116984 Dictyostelium discoideum cl

Query Match

Best Local Similarity 5.1%; Score 73.2; DB 3; Length 110000;

Matches 198; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

Qy 1029 AAAATAAGGTTTATAGGTAAAGCCCAATGTTATTTTGGCATGGAGGCTTTAAA 1088

```
Db 10330 AAAATGATCACCTTCTTTTCCTTTTATAATCGTTTCATTTTTTTTTTAATTTTTTTT 10271
QY 1089 ATTGTGCTCTTTTTCATATTTTATTCATATTCATATTCATATTCATATTCATATTCAT 1148
Db 10270 TTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTT 10211
QY 1149 GGAGATAATATATGTTATATATTTAGTTTGGGGGGAATATTTGTCGCAAGAGGATAATT 1208
Db 10210 CCTGCAAAAAATTTGTGGGCAACCTCTTAGTATTTTAAATTTTAAAAAAGTTCGGTTT 10151
QY 1209 TAAATTACGTCTTCTGTTATTCAGATTAAGAGAGAGACTACGCTGCATATTCAGAG 1268
Db 10150 TAGGTGCGCATTCCTCTCACATGCGAAAAAATAAATAAATAAATAAATAAATAAATAA 10091
QY 1269 TTGTACCTTTAAACATTCGTGAAACATTTTTTCTAAGATTTTCAAAAGGAATATGTAAAT 1328
Db 10090 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 10031
QY 1329 TGAGAAATCATACCACTGCTCTTAACCTTGGTAAACAACTGTTCTTAAATAAAGTATTTA 1388
Db 10030 ATACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 9971
QY 1389 ATGATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1434
Db 9970 ATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 9925
```

Search completed: July 20, 2004, 18:51:00
Job time : 5507 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 15:59:46 ; Search time 583 Seconds
(without alignments)

10449.256 Million cell updates/sec

Title: US-10-799-747-12

Perfect score: 1434

Sequence: 1 cattaactcttttcttcgg.....aaaaaaaaaaaaaaaaaaaaa 1434

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002s:*
- 7: geneseqn2003as:*
- 8: geneseqn2003bs:*
- 9: geneseqn2003cs:*
- 10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1432	99.9	1434	7 ADA39771	Ada39771 Human sec
2	1432	99.9	1434	7 ACC50424	Acc50424 Human sec
3	1430.4	99.7	1434	2 AAX37452	Aax37452 Human sec
C 4	425.8	29.7	448	6 ABN94887	Abn94887 Gene #138
5	360	25.1	415	3 AAH30357	Aah30357 Human col
6	356.8	24.9	395	3 AAC10084	Aac10084 Human sec
7	300.4	20.9	336	2 AAT25136	Aat25136 Human gen
C 8	69.8	4.9	545	4 AAH70126	Aah70126 Human cer
C 9	69	4.8	8056	7 ABZ10246	Abz10246 Haematopo
C 10	67.8	4.7	419	7 ABX46069	Abx46069 Bovine ES
C 11	65.6	4.6	597	4 AAH71472	Aah71472 Human cer
12	63.8	4.4	8056	7 ABZ10246	Abz10246 Haematopo
13	62.6	4.4	8056	7 ABZ10100	Abz10100 Haematopo
C 14	62.6	4.4	8056	7 ABZ10100	Abz10100 Haematopo
15	62	4.3	2103	5 AAD08708	Aad08708 Lycopersi
C 16	61.2	4.3	393	7 ABX39417	Abx39417 Bovine ES
17	61.2	4.3	887	3 AAC59237	Aac59237 Human sec
18	60.8	4.2	2195	3 AAAL4072	Aaal4072 Human SPR
C 19	60.6	4.2	555	3 AAC66747	Aac66747 DNA marke
20	60.4	4.2	329	6 ABN94208	Abn94208 Gene #706
21	60.2	4.2	6076	4 AAS46664	Aas46664 Tumour su
22	60	4.2	6195	6 ABL32590	AbL32590 Human imm
C 23	60	4.2	17389	6 ABL33415	AbL33415 Human imm

C 24	59.8	4.2	6065	6 ABK31357	Abk31357 Signal tr
C 25	59.8	4.2	6065	6 ABL70580	AbL70580 Chemical
C 26	59.8	4.2	6065	6 AAS61261	Aas61261 Human gen
C 27	59.2	4.1	337	7 ABX41811	Abx41811 Bovine ES
28	59.2	4.1	17294	6 ABL32986	AbL32986 Human imm
29	59	4.1	476	8 ACH28349	Ach28349 Human adu
30	59	4.1	7352	6 ABL32370	AbL32370 Human imm
31	58.8	4.1	2976	7 ABZ10131	Abz10131 Haematopo
32	58.8	4.1	3366	4 AAF24684	Aaf24684 Nucleotid
33	58.8	4.1	3366	4 AAE24706	Aaf24706 Nucleotid
34	58.8	4.1	5976	9 ADEB4233	AdB54233 Pretreat
35	58.8	4.1	5976	9 ADEB4171	AdB54171 Human lym
36	58.8	4.1	10365	9 ADD94068	Add94068 Human ATP
37	58.8	4.1	10381	9 ADD94069	Add94069 Human ATP
38	58.8	4.1	10423	9 ADD94070	Add94070 Human ATP
39	58.8	4.1	10442	4 AAF24680	Aaf24680 Nucleotid
40	58.8	4.1	10442	4 AAF24702	Aaf24702 Nucleotid
41	58.8	4.1	10474	4 AAF24685	Aaf24685 Nucleotid
42	58.8	4.1	10474	4 AAF24686	Aaf24686 Nucleotid
43	58.8	4.1	10474	4 AAF24708	Aaf24708 Nucleotid
44	58.8	4.1	10474	4 AAF24707	Aaf24707 Nucleotid
45	58.6	4.1	424	7 ABX46053	Abx46053 Bovine ES

ALIGNMENTS

RESULT 1

ADA39771

ID ADA39771 standard; cDNA; 1434 BP.

XX ADA39771;

AC ADA39771;

XX ADA39771;

DT 20-NOV-2003 (first entry)

XX Human secreted protein encoding cDNA.

DE Human secreted protein encoding cDNA.

XX Human; secreted protein; cancer; hyperproliferative disorder;

KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;

KW anaemia; allergic reaction; asthma; cardiovascular disorder;

KW wound healing; cytostatic; immunosuppressive; neutropenic; neuroprotective;

KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;

KW vulnery; cardiac; gene therapy; ss.

XX Homo sapiens.

OS Homo sapiens.

XX WO2002102993-A2.

PN WO2002102993-A2.

XX 27-DEC-2002.

PD 27-DEC-2002.

XX 19-MAR-2002; 2002WO-US008123.

XX 19-MAR-2002; 2002WO-US008123.

XX 21-MAR-2001; 2001US-0277340P.

PR 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-175238/17.

XX New human secreted proteins and nucleic acid molecules, useful for

XX preparing a diagnostic or pharmaceutical composition for diagnosing,

XX preventing or treating cancer or other hyperproliferative disorder,

XX asthma, allergies or AIDS.

XX Claim 9; SEQ ID NO 153; 3205pp; English.

XX The invention relates to novel genes ADA39629-ADA40565 and proteins

XX ADA40566-ADA41501 for human secreted proteins, useful for preventing,

XX treating or ameliorating medical conditions e.g. by protein or gene

XX therapy. The polypeptides, nucleic acid molecules, antibodies or their

CC fragments, and agonists or antagonists that bind to the polypeptide are
 CC useful for preparing a diagnostic or pharmaceutical composition for
 CC diagnosing or treating cancer or other hyperproliferative disorder. The
 CC polypeptides and nucleic acid molecules are also useful for detecting,
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
 CC or other hyperproliferative disorders including neoplasms, autoimmune
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
 CC thrombocytopenia), allergic reactions including asthma or eczema,
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
 CC fungal or viral infections including HIV/AIDS), or wound healing and
 CC disorders of epithelial cell proliferation. The nucleic acids are also
 CC useful for chromosome identification, radiation hybrid mapping or long-
 CC range restriction mapping, as molecular weight markers, or as
 CC hybridization or diagnostic probes. The polypeptides and antibodies are
 CC useful for providing immunological probes for differential identification
 CC of the tissues immunohistochemistry assays. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pot_sequences.

SQ Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Query Match 99.9%; Score 1432; DB 7; Length 1434;

Best Local Similarity 100.0%; Pred. No. 1.5e-225;

Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATTAACTCTTTTATCGGAATAGTATGATATTTCAATGTCACCTCCATTCAATGTTGA 60
 DB |||||||
 QY 1 CATTAACTCTTTTATCGGAATAGTATGATATTTCAATGTCACCTCCATTCAATGTTGA 60
 DB |||||||
 QY 61 TTTGGAGCTGACAGTATTTTGTGTAAGCAGAGATTTAAATTTATATTGAAAGTCAGTGC 120
 DB |||||||
 QY 61 TTTGGAGCTGACAGTATTTTGTGTAAGCAGAGATTTAAATTTATATTGAAAGTCAGTGC 120
 DB |||||||
 QY 121 AAAATTATGAATAGGATATCTAATAAATCAAAAGTAAATCAAAAGTCAGAGTGT 180
 DB |||||||
 QY 121 AAAATTATGAATAGGATATCTAATAAATCAAAAGTAAATCAAAAGTCAGAGTGT 180
 DB |||||||
 QY 181 CTTAAATAAAATCTCGGTTCTTAAATTTATTTAAATTTATCTTGAATAGTTTCT 240
 DB |||||||
 QY 181 CTTAAATAAAATCTCGGTTCTTAAATTTATTTAAATTTATCTTGAATAGTTTCT 240
 DB |||||||
 QY 241 TAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300
 DB |||||||
 QY 241 TAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300
 DB |||||||
 QY 301 CAAATTTCTATTAATGCAMGAGTGTATATATACAGAAATTTATCGGCATTACCAAGTC 360
 DB |||||||
 QY 301 CAAATTTCTATTAATGCAMGAGTGTATATATACAGAAATTTATCGGCATTACCAAGTC 360
 DB |||||||
 QY 361 TAGGCACATATAGGAATGACGACTCAGAAATGTTTCAATGTAGTATGATGCTGTGTA 420
 DB |||||||
 QY 361 TAGGCACATATAGGAATGACGACTCAGAAATGTTTCAATGTAGTATGATGCTGTGTA 420
 DB |||||||
 QY 421 AGGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTCTAAATGCTGTCTCAATGCTGG 480
 DB |||||||
 QY 421 AGGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTCTAAATGCTGTCTCAATGCTGG 480
 DB |||||||
 QY 481 CCTTTGGCTACTCTACTTCCGATTTATGCGAGCCCATTCAGTCTGAGTTTCTTCTCT 540
 DB |||||||
 QY 481 CCTTTGGCTACTCTACTTCCGATTTATGCGAGCCCATTCAGTCTGAGTTTCTTCTCT 540
 DB |||||||
 QY 541 GGACACCTTATGCTCTGAAATCATCAGCGAGGCTGATTTCAATTTGGTGAATGGGTAGAAA 600
 DB |||||||
 QY 541 GGACACCTTATGCTCTGAAATCATCAGCGAGGCTGATTTCAATTTGGTGAATGGGTAGAAA 600
 DB |||||||
 QY 601 GCAGTATGTTTGTCTGACATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 660
 DB |||||||

DB 601 GCAGTATGTTTGTCTGACATTAAGATGATGATGATGATGATGATGATGATGATGAT 660
 QY |||||||
 DB 661 GTTTTATATCTTTAAATAAGAAATTAACCTTTTAAAGCTATTCCACTCTCTCCCCAGC 720
 QY |||||||
 DB 661 GTTTTATATCTTTAAATAAGAAATTAACCTTTTAAAGCTATTCCACTCTCTCCCCAGC 720
 QY |||||||
 DB 721 CTATCTCAAACTGGTGGATATATGAGAGATCTTGAAGAAGTAAATTAATTAATTAAT 780
 QY |||||||
 DB 721 CTATCTCAAACTGGTGGATATATGAGAGATCTTGAAGAAGTAAATTAATTAATTAAT 780
 QY |||||||
 DB 781 GCTCCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTAATAC 840
 QY |||||||
 DB 781 GCTCCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTAATAC 840
 QY |||||||
 DB 841 TTACCTTCTATTTCTGAAATCAGTTGTGAATCTGTTGCTTATGTTTCAGARGTTTAAAGAC 900
 QY |||||||
 DB 841 TTACCTTCTATTTCTGAAATCAGTTGTGAATCTGTTGCTTATGTTTCAGARGTTTAAAGAC 900
 QY |||||||
 DB 901 CTCMGTAATTCATTTTAAATCTGTAATCTGTAAGACATTTGAATGAATTTCTTAACA 960
 QY |||||||
 DB 901 CTCMGTAATTCATTTTAAATCTGTAATCTGTAAGACATTTGAATGAATTTCTTAACA 960
 QY |||||||
 DB 961 AGAAGACTCATCTGTAGCTGTTTGTGACTCTCTATGAGCCCATTAAGGGTCTGTGCTTA 1020
 QY |||||||
 DB 961 AGAAGACTCATCTGTAGCTGTTTGTGACTCTCTATGAGCCCATTAAGGGTCTGTGCTTA 1020
 QY |||||||
 DB 1021 GCATTAACAAAATAAGSTTTATAGGTAAGCCCAATGTAATTTTTCATGATGAGS 1080
 QY |||||||
 DB 1021 GCATTAACAAAATAAGSTTTATAGGTAAGCCCAATGTAATTTTTCATGATGAGS 1080
 QY |||||||
 DB 1081 GCTTTAAATTTCTGCTCTTTTCAATTTTATTTTCAATTTTATGTTTGTAACTG 1140
 QY |||||||
 DB 1081 GCTTTAAATTTCTGCTCTTTTCAATTTTATTTTCAATTTTATGTTTGTAACTG 1140
 QY |||||||
 DB 1141 CTTTGTAGGAGATATATATATGTTTAAATTTAGTTTGGGGGGAATAATTTGTGCAAGA 1200
 QY |||||||
 DB 1141 CTTTGTAGGAGATATATATATGTTTAAATTTAGTTTGGGGGGAATAATTTGTGCAAGA 1200
 QY |||||||
 DB 1201 GGATTAATTTTAAATTTAGTCTCTGTTTATTTTCAAGATTAAGAGAGACACTACGCTGATA 1260
 QY |||||||
 DB 1201 GGATTAATTTTAAATTTAGTCTCTGTTTATTTTCAAGATTAAGAGAGACACTACGCTGATA 1260
 QY |||||||
 DB 1261 TTCAAGAGTTGTACCTTAACATTTGTTGAAACATTTTCTAAGATTTTCAAAAGGAATAT 1320
 QY |||||||
 DB 1261 TTCAAGAGTTGTACCTTAACATTTGTTGAAACATTTTCTAAGATTTTCAAAAGGAATAT 1320
 QY |||||||
 DB 1321 GTGTAAATTCAGAAATCATACCACTGCTCTAACTTGGTAAACAAACTGTTCTTAAATAA 1380
 QY |||||||
 DB 1321 GTGTAAATTCAGAAATCATACCACTGCTCTAACTTGGTAAACAAACTGTTCTTAAATAA 1380
 QY |||||||
 DB 1381 AGTATTTAATGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1434
 QY |||||||
 DB 1381 AGTATTTAATGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1434
 QY |||||||

RESULT 2

ACC50424

ID ACC50424 standard; cdna; 1434 BP.

XX ACC50424;

XX AC

XX 12-JUN-2003 (first entry)

XX Human secreted protein coding sequence, SEQ ID 91.

XX Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;

XX vulnery; antiinflammatory; nootropic; neuroprotective;

XX antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.

XX Homo sapiens.

XX OS

XX PN

XX WO200295010-A2

XX

RESULT 3

AA37452
ID AAX37452 standard; cDNA; 1434 BP.

XX AC AAX37452;
XX DT 06-JUL-1999 (first entry)
XX DE Human secreted protein cDNA fragment containing gene 2.

XX Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.

XX OS Homo sapiens.

XX PN W09918208-A1.

XX PD 15-APR-1999.

XX PF 01-OCT-1998; 98WO-US020775.

XX PR 02-OCT-1997; 97US-0060833P.

XX PR 02-OCT-1997; 97US-0060836P.

XX PR 02-OCT-1997; 97US-0060837P.

XX PR 02-OCT-1997; 97US-0060838P.

XX PR 02-OCT-1997; 97US-0060839P.

XX PR 02-OCT-1997; 97US-0060843P.

XX PR 02-OCT-1997; 97US-0060862P.

XX PR 02-OCT-1997; 97US-0060866P.

XX PR 02-OCT-1997; 97US-0060874P.

XX PR 02-OCT-1997; 97US-0060880P.

XX PR 02-OCT-1997; 97US-0060884P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;

XX PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Peng P;

XX PI Lafleur DW, Shi Y;

XX WPI; 1999-264022/22.

XX P-PSDB; AAY07853.

XX New isolated human genes and the secreted polypeptides they encode.

XX Claim 1a; Page 228; 368pp; English.

XX This invention describes novel isolated human genes and the secreted

XX proteins they encode. The products of the invention are useful for

XX preventing, treating or ameliorating medical conditions, e.g. by protein

XX or gene therapy. Also pathological conditions can be diagnosed by

XX determining the amount of the new polypeptides in a sample or by

XX determining the presence of mutations in the new polynucleotides.

XX Specific uses are described for each of the 101 polynucleotides, based on

XX which tissues they are most highly expressed in, and include developing

XX products for the diagnosis or treatment of cancer, tumours,

XX neurodegenerative disorders, developmental abnormalities and fetal

XX deficiencies, blood disorders, leukemias, diseases of the immune system,

XX autoimmune diseases, hepatic and renal diseases, lymphomas, inflammation,

XX allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate

XX disease, skeletal or cardiac muscle disorders, pulmonary disorders,

XX transplant rejection, disorders involving osteoclasts such as

XX osteoporosis, arthritis or malignancies, digestive/endocrine disorders,

XX infections and AIDS. The human secreted proteins of the invention are

XX represented in AAY07852-Y07993 and the encoding nucleic acids are

XX represented in AAX37451-X37552

SQ Sequence 1434 BP; 480 A; 204 C; 250 G; 495 T; 0 U; 5 Other;
Query Match 99.7%; Score 1430.4; DB 2; Length 1434;
Best Local Similarity 99.9%; Pred. No. 2.7e-225;
Matches 1433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CATTAAACTCTTTTATCGGGAATAGTATGATATATTTCAATGTGCACATCCATTCATGTGA 60
DB 1 CATTAAACTCTTTTATCGGGAATAGTATGATATATTTCAATGTGCACATCCATTCATGTGA 60
QY 61 TTTGGAGCTGACAGTATTTTGTGTAGCAGAGATTTAAATTTTATATTTGAAGTCAGTGC 120
DB 61 TTTGGAGCTGACAGTATTTTGTGTAGCAGAGATTTAAATTTTATATTTGAAGTCAGTGC 120
QY 121 AAAATTATGATAGGATATCTAATAATAACAAGTAAATAACAAGTCAAGACAGTGT 180
DB 121 AAAATTATGATAGGATATCTAATAATAACAAGTAAATAACAAGTCAAGACAGTGT 180
QY 181 CTAATAAAAAATTTCTGGGTTCTTAAATAATTTTAAATTTATCTTGAATAGTTTCT 240
DB 181 CTAATAAAAAATTTCTGGGTTCTTAAATAATTTTAAATTTATCTTGAATAGTTTCT 240
QY 241 TAGATTAACTCTCAGGATATGAGAAAGTCAATTAAGTGTGATAGTAAAGTATCATPAAA 300
DB 241 TAGATTAACTCTCAGGATATGAGAAAGTCAATTAAGTGTGATAGTAAAGTATCATPAAA 300
QY 301 CAAATTGTCTATTAAATGCMAGAGTGGTAATATACAGAAATTTATCAGGCATTTACCAAGTC 360
DB 301 CAAATTGTCTATTAAATGCMAGAGTGGTAATATACAGAAATTTATCAGGCATTTACCAAGTC 360
QY 361 TAGGCACATATAGGAATGACAGCACTCAGAAATGGTTTCAATGTAGTAGTGTGCTTGA 420
DB 361 TAGGCACATATAGGAATGACAGCACTCAGAAATGGTTTCAATGTAGTAGTGTGCTTGA 420
QY 421 AGGTAGGGAGCTTTATTCAGACATAGTAGATAGTTCTCTPAATGCTGTSTCAATTTGCTG 480
DB 421 AGGTAGGGAGCTTTATTCAGACATAGTAGATAGTTCTCTCTAATGCTGTSTCAATTTGCTG 480
QY 481 CTTTGGCTACCTGTACTTCCSCATTTATGCGACCCCATTCAGTCTTGAGTTTTCTTCTCT 540
DB 481 CTTTGGCTACCTGTACTTCCSCATTTATGCGACCCCATTCAGTCTTGAGTTTTCTTCTCT 540
QY 541 GGACACCTTATGCTCTGAAATCATGACGAGGCTGATTCATTTGGTGATTTGGGTAGAAA 600
DB 541 GGACACCTTATGCTCTGAAATCATGACGAGGCTGATTCATTTGGTGATTTGGGTAGAAA 600
QY 601 GCAGTATGTTTTGCTGACATTAAGATGTAGGTATAGATAGTGGTTTAAAGTGTAT 660
DB 601 GCAGTATGTTTTGCTGACATTAAGATGTAGGTATAGATAGTGGTTTAAAGTGTAT 660
QY 661 GTTTTTATCTTTAAATAAAGAAATATACTTTTAAAGCTATTTCACCTCTCCCCGAGC 720
DB 661 GTTTTTATCTTTAAATAAAGAAATATACTTTTAAAGCTATTTCACCTCTCCCCGAGC 720
QY 721 CTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAATCACTTCACT 780
DB 721 CTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAATCACTTCACT 780
QY 781 GCTCCACTCCAGTGAATCGCCCACTCCCACTGACCTAGTAGAAATTTGAATTTAATAC 840
DB 781 GCTCCACTCCAGTGAATCGCCCACTCCCACTGACCTAGTAGAAATTTGAATTTAATAC 840
QY 841 TTACCTTCTATTCTCAAAATCACTGTGAACTGTGCTTATCTGCTCAGAGTTTAAAGAAC 900
DB 841 TTACCTTCTATTCTGAAATCACTGTGAACTGTGCTTATCTGCTCAGAGTTTAAAGAAC 900
QY 901 CTCMGTCATTTCAATTTTAAATCTGCTATTCTGAGAAGCATTTGAATGAATTTCTTAACA 960
DB 901 CTCMGTCATTTCAATTTTAAATCTGCTATTCTGAGAAGCATTTGAATGAATTTCTTAACA 960
QY 961 AGAAGACTCATCTGAGTCTTTGCTGACTCTCTGAGCCCATAGGGTCTGTGCTTA 1020
DB 961 AGAAGACTCATCTGAGTCTTTGCTGACTCTCTGAGCCCATAGGGTCTGTGCTTA 1020


```
PR 27-OCT-1998; 98US-0105877P.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
XX Lamson G, Drmanac R, Kravzenjakov R, Dickson M, Drmanac S, Labat I;
XX Leeshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
XX WPI; 2000-293155/25.
XX
XX Polynucleotide library comprising 1079 defined sequences, useful in the
XX form of an array to detect cancer or susceptibility to cancer.
XX
XX Claim 1; Page 268; 502pp; English.
XX
XX The present invention describes a library of polynucleotides comprising
XX 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
XX are: (1) an isolated polynucleotide (I) having at least 90% identity to
XX one of the 1079 sequences; (2) a recombinant host cell containing (I);
XX (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
XX specifically binds to (II); (5) a vector comprising (I); and (6) a method
XX of detecting differentially expressed genes correlated with a cancerous
XX state of a mammalian cell comprising detecting a gene product encoded by
XX 65 of the 1079 sequences given in the specification. The polynucleotides
XX are used to monitor patients having (or susceptible) to cancer to detect
XX potentially malignant events at a molecular level before they are
XX detectable at a gross morphological level. The polynucleotides are also
XX useful for monitoring the efficacy of various therapies and preventive
XX interventions. Polynucleotide probes based on the disclosed sequences are
XX useful for chromosome mapping and detection of transcription levels. The
XX 1079 polynucleotide sequences were derived from a human colon cancer cell
XX line Km12L4-A cDNA library
XX
XX Sequence 415 BP; 114 A; 73 C; 87 G; 141 T; 0 U; 0 Other;
XX
XX Query Match 25.1%; Score 360; DB 3; Length 415;
XX Best Local Similarity 94.4%; Pred. No. 4e-50;
XX Matches 391; Conservative 3; Mismatches 18; Indels 2; Gaps 2;
XX
XX 302 AAATTGCTATTAAATGAMGA-GTGTAATATACAGAAATTTATCAGGCAATACCAAGTC 360
XX 1 AAATTGCTATTAAATGAMGAATGAGTGTGTAATATACAGAAATTTATCAGGCAATACCAAGTC 60
XX
XX 361 TAGGCACATATAGGAATGAGCACTCAGAAATGTTTCAATGTAGTGTGATGCTGTGA 420
XX 61 TAGGCACATATAGGAATGAGCACTCAGAAATGTTTCAATGTAGTGTGATGCTGTGA 120
XX
XX 421 AGGTAGGGAGCTTATTTCAGACATAGTAGATAGTCTCTTAATGCTGTCTCAATTTGCTGG 480
XX 121 AGGTAGGGAGCTTATTTCAGACATAGTAGATAGTCTCTTAATGCTGTCTCAATTTGCTGG 180
XX
XX 481 CTTTGGCTACCTGTACTTCGCAATATGAGGAGGCTGATTTCAATTTGCTGTGATTTCTCTCT 540
XX 181 CTTTGGCTACCTGTACTTCGCAATATGAGGAGGCTGATTTCAATTTGCTGTGATTTCTCTCT 239
XX
XX 541 GGACACCTTATGCTCTGAATATCATGAGGAGGCTGATTTCAATTTGCTGTGATTTGCTGAAA 600
XX 240 GGACACCTTATGCTCTGAATATCATGAGGAGGCTGATTTCAATTTGCTGTGATTTGCTGAAA 299
XX
XX 601 GCAGTATGTTTCTGACATTAAGATAGTAGGTTATAGATAGGTTTACGCTTTTAAGTGTAT 660
XX 300 GCAGTATGTTTCTGACATTAAGATAGTAGGTTATAGATAGGTTTACGCTTTTAAGTGTAT 359
XX
XX 661 GTTTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGCTTATTCCTCCTCC 714
XX 360 GTTTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGCTTATTCCTCCTCC 413
XX
XX RESULT 6
XX AAC10084
XX ID AAC10084 standard; cDNA; 395 BP.
```

```
XX AAC10084;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 14159.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EPI033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.
XX
XX 26-FEB-1999; 99US-0122487P.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 1; SEQ ID NO 14159; 71pp + Sequence Listing; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX
XX Sequence 395 BP; 147 A; 41 C; 74 G; 125 T; 0 U; 8 Other;
XX
XX Query Match 24.9%; Score 356.8; DB 3; Length 395;
XX Best Local Similarity 95.7%; Pred. No. 1.3e-49;
XX Matches 378; Conservative 8; Mismatches 6; Indels 3; Gaps 2;
XX
XX 58 TGATTTGGAGCTGACAGTATTATTTGTGTAGCAGAGATTTAATTTATATTTGAAGTCAG 117
XX 1 TGATTTGGAGCTGACAGTATTATTTGTGTAGCAGAGATTTAATTTATATTTGAAGTCAG 60
XX
XX 118 TGCAGAAATTATGATAGGATATATTAATAACAAGTAATAACAAGTCAAGCAAGT 177
XX 61 TGCAGAAATTATGATAGGATATATTAATAACAAGTAATAACAAGTCAAGCAAGT 120
XX
XX 178 GTTCTAAATAAATAATCTGGTTCCTTAAATAATTTTAAATTT--ATCTTCAAAATAGT 235
XX 121 GTTCTAAATAAATAATCTGGTTCCTTAAATAATTTTAAATTTTATCTTTGAAATAGT 180
XX
XX 236 TTTCTTAGATTAATCTCAGATATGAGAAAGTCAATTAAGTGTGAGTAAGTAGTATCA 295
XX 181 NMTCTTAGATTAATCTCAGATATGAGAAAGTCAATTAAGTGTGAGTAAGTAGTATCA 240
XX
XX 296 TTAACAAAATTTGCTATTAAATGTCAMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 354
XX 241 TTAACAAAATTTGCTATTAAATGTCAMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 300
```

Db	61	CACTGACCTAGTAGAATTTGTAAATTTAATACTTTACCTTCTATTCTGAATCAGTTGTGA	120
Qy	870	ACTGTTGCCTTATGTTTCAGAGGTTTAAAGAACCTCMGTGAATTCATTTTAAAAATCTGCT	929
Db	121	ACGTTTGCCTTAAGTTTCAGAGGTTTAAAGAACCTCAGTGAATTCATTTTAAAAATCTGCT	180
Qy	930	ATTCTGAGAAGCAATCAATGAATTTCTTAAACAAGAGCACTCATCTGTAGCTGTGTTGCTGAC	989
Db	181	ATTCTGAGAAGCAATCAATGAATTTCTTAAACAAGAGCACTCATCTGTAGCTGTGTTGCTGAC	240
Qy	990	TCCATGAGCCCCATTAAGGTTCTGTGCTTAGCATTAACAATAAGGTTTATAGG--TA	1047
Db	241	TCCATGAG--CCATTAAGGTTCTGTGCTTAGCATTAACAATAAGGTTTATAGGTTAA	299
Qy	1048	AAGCCAATGTATTAATTTTTTTTTTTCATGAGGGC	1082
Db	300	AAGCCAATGTATTAATTTTTTTTTTTCATGAGGGC	334
RESULT 8			
ID	AAH70126/c		
ID	AAH70126 standard; cDNA; 545 BP.		
AC	AAH70126;		
XX			
DT	19-SEP-2001 (first entry)		
XX			
DE	Human cervical cancer marker nucleic acid 1400.		
XX			
KW	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.		
OS	Homo sapiens.		
XX			
PN	W0200142467-A2.		
XX			
PD	14-JUN-2001.		
XX			
XX	08-DEC-2000; 2000WO-US033312.		
PR	08-DEC-1999; 99US-0169681P.		
PR	21-DEC-1999; 99US-0171350P.		
PR	14-MAR-2000; 2000US-0189315P.		
PR	12-MAY-2000; 2000US-0203791P.		
PR	09-JUN-2000; 2000US-0210600P.		
PR	21-JUL-2000; 2000US-0220114P.		
XX			
PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.		
XX			
PI	Schlegel R, Deeds J, Berger A, Zhao X;		
XX	WPI; 2001-375006/39.		
DR			
XX			
PT	New isolated nucleic acid for diagnosing and treating cervical cancer and		
PT	for assessing and detecting compounds for treating the cancer.		
XX			
PS	Claim 1; Page 319-320; 1051pp; English.		
XX			
CC	The invention relates to novel genes (AAH68727-AAH73393) associated with		
CC	cervical cancer with cytostatic activity. The nucleic acids and encoded		
CC	polypeptides are useful: to assess if a patient is afflicted with		
CC	cervical cancer or has a pre-malignant condition; to monitor the		
CC	progression of cervical cancer or a premalignant condition in a patient;		
CC	and to select and/or assess the efficacy of a compound or therapy for		
CC	inhibiting cervical cancer in a patient. The nucleic acids may also be		
CC	useful for gene therapy		
XX			
SQ	Sequence 545 BP; 200 A; 40 C; 23 G; 209 T; 0 U; 73 Other;		
Query Match 4.9%; Score 69.8; DB 4; Length 545;			
Best Local Similarity 43.9%; Pred. No. 0.013;			
Matches 167; Conservative 0; Mismatches 213; Indels 0; Gaps 0;			
Qy	1055	TGTATTAATTTTTTTTTTTCATGAGGGCTTTAAAAATTTGCTCTTTTCATATTATT	1114

Db 410 TTTTAAANNTTTNNNNNTTTTAAAAAANNTTTTNTTNAATTTTTTTA 351
 QY 1115 CATATTCAATTTATGTTTGTAACTGCTTTTAGGAGATAATTATATGTTATAAATTAG 1174
 Db 350 AATTGGAATTTTAAAAANTANTGNTNAACNNGNCRAAAAATNAANTTTTTTNTT 291
 QY 1175 TTTTGGGGGAATTAATGTGCAAGAGATAATTTAATTTACGTGCTTCTGTTATTCAGA 1234
 Db 290 TTTTNAAAAAANNTTTTNCCTNNNANTTTNNNATNTTTTTTTTTTGNANTTTTTTCCC 231
 QY 1235 ATAAAGAGAGACGTACGCTGCATATTCAGAGTTGTACCTTAAACATTCGTGCAACATT 1294
 Db 230 CCCAATTTTTTTTTNNNNNTTTTTTNNNNNTTTTNNNTTTTNNNTTTTNNAAAAAANNTT 171
 QY 1295 TTTTCTAAGATTTTCAAAAGGAATATGTGTAAATTTGAGAAATCATACCACTGTCCTAAC 1354
 Db 170 TTTTNAANTTTTTTNAAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAA 111
 QY 1355 TTGGTAAACAACTGTTCTTAATAAAGTATTTAATGATTTTAAAAAANNTTTTAAAAA 1414
 Db 110 AAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAA 51
 QY 1415 AAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAA 1434
 Db 50 GAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAA 31

RESULT 9
 ABZ10246/c
 ID ABZ10246 standard; DNA; 8056 BP.
 XX
 AC ABZ10246;
 XX
 DT 16-JAN-2003 (first entry)
 DE Haematopoietic cell proliferation disorder related DNA sequence #386.
 XX
 KW Human, haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 XX cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO20027272-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-EP003401.
 XX
 PR 26-MAR-2001; 2001US-0278333P.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
 PI Schwoppe I, Ziebarth H;
 XX
 DR WPI; 2003-018942/01.
 XX
 XX
 PT Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent that
 PT distinguishes between methylated and non-methylated CpG dinucleotides.
 XX
 XX
 PS Claim 28; SEQ ID NO 386; 117pp; English.
 XX
 XX The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG

CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclases, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;
 Query Match 4.8%; Score 69; DB 7; Length 8056;
 Best Local Similarity 48.0%; Pred. No. 0.017;
 Matches 293; Conservative 1; Mismatches 306; Indels 11; Gaps 3;
 QY 826 TTGTAAATTAATACCTTCTCTATTCTGAAATCAGTTGTGAACCTGTTCCTATGTT 885
 Db 2006 TTATTAAATTTTAAATAAATAATTAATTAATTTTATTAAAAATATATTTTTT 1947
 QY 886 CAGAGTTTAAAGACCTCMGTGAATTCATTTTTTAAAACTGCTATTCGAGAGCAATG 945
 Db 1946 TATTATATTAATTAATTAATTTTATTTTATATAAATTTTATTTTATATAAAAA 1887
 QY 946 AATGAATTCCTTACAGAGACCTCATCTGTAGCTGTTTGTGACTCCTATGAGCCCAT 1005
 Db 1886 AATAAAAAATAATTAATTAATTTTATTTTATTTTATTTTATTTTAAAAACAAA 1827
 QY 1006 AGGGTCTGTGCTTAGCAATTA---ACAAATAAGGTTTAGTGAAGCAATGTATTA 1061
 Db 1826 TTTTATTATTATTAAATTAATAATTAATAATTAATTAATTTTAAATTTTAAATTA 1767
 QY 1062 ATTTTTTTTGCATGAGGGCTTTAAAAATTTGCTCTTTTTCATATTTTATTCATATTC 1121
 Db 1766 AAAAAATTTTTTAAATTTTAAATAATTTAAATAATTAATAATATATATATATAT 1707
 QY 1122 AATTTATGTTTGTAACTGCTTTTAGGGAGATAATTATATGTATATAAATTAGTTTGG 1181
 Db 1706 AAATTAATTTTATAAT--TTAATAAAAAATAAATTTTATCATATTAATTAATAAT 1649
 QY 1182 GGGATAAATTTGTCAAGAGGATAATTTAATTTACGTCCTTCTGTTATTCAGATAAAGA 1241
 Db 1648 TTTTAAATTTTACAA---TTCAATTTATTTCAATTTTAAAAATATTAATAA 1594
 QY 1242 GAGAAGACTACGCTGCATATTCAGAGTTGTACCTTAAACATTCGTTGNAACATTTTCTA 1301
 Db 1593 TATAAATTAATAATTAATAATTTAAATAATTTTAAAAAATAAATTTTATTTAAAAAT 1534
 QY 1302 AGATTTTCAAAAGGATATGTTAAATTCAGAAATCATACCACTGTCCTAACTGTGTA 1361
 Db 1533 TTTTAAACATATTTATTTTATATTTATATATATATATATTAATTAATTAATTTTAA 1474
 QY 1362 ACAAACTGTTCTTTAAATAAAGTATTTTAATGATTTTAAAAAANNTTTTAAAAA 1421
 Db 1473 ATATAATTTTATTAATAAATAAATTTTATTTTATTTTAAAAAANNTTTTAAAAA 1414
 QY 1422 AAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAA 1432
 Db 1413 TAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAAANNTTTTAAAAA 1403
 RESULT 10
 ABX46069/c
 ID ABX46069 standard; cDNA; 419 BP.
 XX
 AC ABX46069;

ABZ10100/c
 ID ABZ10100 standard; DNA; 8056 BP.
 XX
 AC ABZ10100;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #240.
 XX
 KW Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FN W0200277272-A2.
 XX
 PD 03-OCT-2002.
 XX
 XX 26-MAR-2002; 2002WO-EP003401.
 PF
 XX 26-MAR-2001; 2001US-0278333P.
 PR
 XX (EPIC-) EPIGENOMICS AG.
 PA
 XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorian P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
 PI Schwöbe I, Ziebarth H;
 XX WPI; 2003-018942/01.
 DR
 XX
 XX
 PT Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent that
 PT distinguishes between methylated and non-methylated CpG dinucleotides.
 XX
 XX Claim 28; SEQ ID NO 240; 117pp; English.
 PS
 XX
 XX The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;
 Query Match 4.4%; Score 62.6; DB 7; Length 8056;
 Best Local Similarity 47.3%; Pred. No. 0.19;
 Matches 289; Conservative 1; Mismatches 310; Indels 11; Gaps 3;
 QY 826 TTGTAAATTAATTAACCTTACCTTATTCGAAATCAGTTGCACTGTTGCTTATGTT 885
 Db
 2006 TTATTAATTTTAAATAAAAATTAATTAATTAATTTTAAATTTTAAATTTTATTTT 1947
 QY 886 CAGAGTTTAAAGACCTCMGTGATTCATTTTAAATCTGCTATTCAGAGCATTC 945

Db 1946 TATTTATATTAATTAATTAATTTTATTTTATTAATAAATTTTATTTTATTAATAA 1897
 QY 946 AATGAATTCCTTAACAAGAGACTCATCTGTAGCTGTTGCTGACTCTCTATGAGCCCAT 1005
 Db 1886 AAATAAAAAATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTAACGAAA 1827
 QY 1006 AGGGTCTGCTGCTTAGCATTA---ACAAAATAAGGTTTATAGTAAAGCCAATGATTA 1061
 Db 1826 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTTTATTTTATTA 1767
 QY 1062 ATTTTTTTTGCATGGAGGGCTTTAAAAATTTGTGCTCTTTTTCATATTTTATTCATATTC 1121
 Db 1766 AAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 1707
 QY 1122 AATTATGTTTGTAACTGCTTTTGTAGGAGAGATAATTAATTAATTAATTAATTAATTAAT 1181
 Db 1706 AAATTAATTTTATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1649
 QY 1182 GCGAATAATTTGTCAAGAGAGATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1241
 Db 1648 TTTTATTAATTTTACGATTCGTTTATTTTCGATTT---TTTAAATATTAATTAATTA 1594
 QY 1242 GAGAGACTACGCTGCATATTTCAAGAGTTGTACCTTAAACATTTGGTGAACATTTTCTA 1301
 Db 1593 TATAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1534
 QY 1302 AGATTTTCAAAAGGATATGTGTAAATTTGAGAATTCATAACCACTGCTTAACTTGGTAA 1361
 Db 1533 TTTTATTAACGATTTTATATTTATATTTATATTAATTAATTAATTAATTAATTAATTT 1474
 QY 1362 ACAAACTGTTCTTAAATAAAGTATTTAAATGATTTTAAATAAATAAATAAATAAATAA 1421
 Db 1473 ATATAATTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1414
 QY 1422 AAAAAAATAA 1432
 Db 1413 TAAAAAATAA 1403
 RESULT 15
 AAD08708
 ID AAD08708 standard; cDNA; 2103 BP.
 XX
 AC AAD08708;
 XX
 DT 04-SEP-2001 (first entry)
 XX
 DE Lycopersicon pennellii elin7 invertase isoenzyme cDNA.
 XX
 KW Invertase isoenzyme; apoplast; brix value; monosaccharide; sugar; fruit;
 KW green-fruited tomato; elin7; ss.
 XX
 OS Lycopersicon pennellii.
 XX
 FH Key Location/Qualifiers
 FT 82..1833
 CDS /tag= a
 FT /product= "elin7 invertase isoenzyme"
 XX
 PN W0200149826-A1.
 XX
 PD 12-JUL-2001.
 XX
 PF 02-JAN-2001; 2001WO-IL000008.
 XX
 PR 04-JAN-2000; 2000US-00477375.
 XX
 PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX
 PI Zamir D, Pleban T, Fridman E;
 XX WPI; 2001-418359/44.
 DR P-PSDB; AAE04416.
 DR

This Page Blank (uspto)

1	59	4.1	832	4	US-09-621-976-2813	Sequence 2813, Ap
2	58.2	4.1	339	4	US-09-621-976-16015	Sequence 16015, A
3	57.6	4.0	189	4	US-09-621-976-14761	Sequence 14761, A
4	56.4	3.9	2887	4	US-08-983-502-14	Sequence 14, Appl
5	56.4	3.9	2887	4	US-09-516-747-14	Sequence 14, Appl
6	56.4	3.9	2887	5	PCT-US96-10521-14	Sequence 14, Appl
7	55.8	3.9	832	4	US-09-621-976-2813	Sequence 2813, Ap
8	55.8	3.9	1172	1	US-07-945-288-9	Sequence 9, Appl
9	55.8	3.9	1172	1	US-08-462-831-9	Sequence 9, Appl
10	55.8	3.9	1172	1	US-08-461-809-9	Sequence 9, Appl
11	55.8	3.9	1172	1	US-08-461-441-9	Sequence 9, Appl
12	55.8	3.9	1172	5	PCT-US93-08518-9	Sequence 9, Appl
13	55.6	3.9	335	4	US-09-621-976-16038	Sequence 16038, A
14	55.4	3.9	19124	2	US-08-487-8268-13	Sequence 13, Appl
15	55	3.8	271	4	US-09-621-976-10193	Sequence 10193, A
16	55	3.8	279	4	US-09-621-976-10220	Sequence 10220, A
17	55	3.8	299	4	US-09-621-976-10211	Sequence 10211, A
18	55	3.8	458	1	US-08-524-757-1	Sequence 1, Appl
19	54.8	3.8	240	1	US-08-628-417-6	Sequence 6, Appl
20	54.8	3.8	335	4	US-09-621-976-16061	Sequence 16061, A
21	54.8	3.8	338	4	US-09-621-976-16041	Sequence 16041, A
22	54.6	3.8	326	4	US-09-621-976-16024	Sequence 16024, A
23	54.6	3.8	327	4	US-09-621-976-16018	Sequence 16018, A
24	54.6	3.8	329	4	US-09-621-976-16012	Sequence 16012, A
25	54.6	3.8	332	4	US-09-621-976-16050	Sequence 16050, A
26	54.6	3.8	332	4	US-09-621-976-16012	Sequence 16053, A
27	54.6	3.8	333	4	US-09-621-976-16032	Sequence 16032, A

Db 2756 TATTTTAAATCATAGGAATTAAGTTATCTTTTAAATTTTAAAGTATCTTTTCCAAAACA 2815
QY 1365 AACTGTTCTTAATAAAGTATTTAATGATTTTAAATAAAAAAAAAAAAAAAAAAAAAA 1424
Db 2816 TTTTAAATAGAAATAAATAATATTGATCTTTAAATAAAAAAAAAAAAAAAAAAAAAA 2875
QY 1425 AAAAAAAAAA 1434
Db 2876 AAAAAAAAAA 2885

RESULT 5
US-09-516-747-14
; Sequence 14, Application US/09516747
; Patent No. 6586571
; GENERAL INFORMATION:
; APPLICANT: David WALLACH
; Mark P. BOLDIN
; Tanya M. GONCHAROV
; Yuri V. GOLTSEV
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
; AND OTHER PROTEINS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/516,747
; FILING DATE: 01-Mar-2000

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/983,502
; FILING DATE: <Unknown>
; APPLICATION NUMBER: IL 114,615
; FILING DATE: 16-JUL-1995
; APPLICATION NUMBER: IL 114,986
; FILING DATE: 17-AUG-1995
; APPLICATION NUMBER: IL 115,319
; FILING DATE: 14-SEP-1995
; APPLICATION NUMBER: IL 116,588
; FILING DATE: 27-DEC-1995
; APPLICATION NUMBER: IL 117,932
; FILING DATE: 16-APR-1996

ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: WALLACH-19
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Query Match 3.9%; Score 56.4; DB 4; Length 2887;
Best Local Similarity 64.6%; Pred. No. 0.0078;
Matches 84; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1305 TTTTCAAAGGAATATGTGTAATTCAGAAATCATAAACCACTGTCTAACTTGGTAAACA 1364
Db 2756 TATTTTAAATCATAGGAATTAAGTTATCTTTTAAATTTTAAAGTATCTTTTCCAAAACA 2815
QY 1365 AACTGTTCTTAATAAAGTATTTAATGATTTTAAATAAAAAAAAAAAAAAAAAAAAAA 1424
Db 2816 TTTTAAATAGAAATAAATAATATTGATCTTTAAATAAAAAAAAAAAAAAAAAAAAAA 2875
QY 1425 AAAAAAAAAA 1434
Db 2876 AAAAAAAAAA 2885

RESULT 6
PCT-US96-10521-14
; Sequence 14, Application PC/TUS9610521
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
; AND OTHER PROTEINS
; NUMBER OF SEQUENCES: 34
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10521
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,615
; FILING DATE: 16-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,986
; FILING DATE: 17-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 115,319
; FILING DATE: 14-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 116,588
; FILING DATE: 27-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 117,932
; FILING DATE: 16-APR-1996
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US96-10521-14

Query Match 3.9%; Score 56.4; DB 5; Length 2887;
Best Local Similarity 64.6%; Pred. No. 0.0078;
Matches 84; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1305 TTTTCAAAGGAATATGTGTAATTCAGAAATCATAAACCACTGTCTAACTTGGTAAACA 1364
Db 2756 TATTTTAAATCATAGGAATTAAGTTATCTTTTAAATTTTAAAGTATCTTTTCCAAAACA 2815
QY 1365 AACTGTTCTTAATAAAGTATTTAATGATTTTAAATAAAAAAAAAAAAAAAAAAAAAA 1424
Db 2816 TTTTAAATAGAAATAAATAATATTGATCTTTAAATAAAAAAAAAAAAAAAAAAAAAA 2875
QY 1425 AAAAAAAAAA 1434
Db 2876 AAAAAAAAAA 2885

RESULT 7
US-09-621-976-2813/c

```
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match          3.9%; Score 55.8; DB 4; Length 832;
Best Local Similarity 11.2%; Pred. No. 0.0083;
Matches 39; Conservative 169; Mismatches 141; Indels 0; Gaps 0;

QY 1085 TAAATTTGTCTCTTTTTCATATTTTATTCATATTCATATTTATGTTGTTGTAAGTCTT 1144
DB 373 TTATATTTTGTCTCTTTTTCATATTTTATTCATATTCATATTTATGTTGTTGTAAGTCTT 1144
QY 1145 TTAGGAGATAATATATATATATATATATATATATATATATATATATATATATATATAT 1204
DB 313 GNNWYMMWGRWSTWYCMYKWKCMYGRRCWYTWARGRWMSYANGKWSMSMSMCT 254
QY 1205 AATTAATTTACGTCTCTGTTATTCAGATTAAGAGAGAACTACGTCGATATCA 1264
DB 253 RMYKKGSTYMKCTCATCYWYKWKEMSKTOWSGRGGYMTSYSTRYSYMWAS 194
QY 1265 AGAGTTGTACCTTAACATTTGTTGAACATTTTCTTAAGATTTTCAAGAGGAATATGTT 1324
DB 193 WMYTMCWGWGRWSTWYMWAGKWKWRYATTWRMMWMMWMMWMMWMMWMMWMMWMMW 134
QY 1325 AATAGAGAAATCATACCACTGCTCTAACTGTTAAACAACTGTTCTTTAAATAAAGTA 1384
DB 133 AMYRRTMMWGYRYWYWKSYRRTRCAWAYAWKTKRSYWCWKCWKCWMMWMMWMMWMMW 74
QY 1385 TTAAATGATTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1433
DB 73 KTMWRACWKIRYRWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMM 25

RESULT 8
US-07-945-288-9
; Sequence 9, Application US/07945288
; Patent No. 5433948
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: Chua, Kaw-Yan
; TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES (HOUSE DUST MITES)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655

; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655

; APPLICATION NUMBER: US/07/945,288
; FILING DATE: 19920910
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: P36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..738
; US-07-945-288-9

Query Match          3.9%; Score 55.8; DB 1; Length 1172;
Best Local Similarity 61.2%; Pred. No. 0.0086;
Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1288 AATCAATTTTCTAAGATTTTCAAGAGGATATGTTAAATTCAGAAATCATACCACTG 1347
DB 1017 AAAGACAATTTCTATATGATTTCTCACTAATTTATTTAAATCAAAATTTTGAATAAG 1076
QY 1348 TCCTAACTTGGTAAACAACTGTTCTTAAATAAGTATTTAAATGATTTTAAATGATTT 1407
DB 1077 AATAAATTCATTCACAAATTTAAATAAATAAATAAATAAATAAATAAATAAATAA 1136
QY 1408 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1434
DB 1137 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1163

RESULT 9
US-08-462-831-9
; Sequence 9, Application US/08462831
; Patent No. 5552142
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,831
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
```

```
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; LENGTH: 1172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..738
; US-08-462-831-9

Query Match          3.9%; Score 55.8; DB 1; Length 1172;
Best Local Similarity 61.2%; Pred. No. 0.0088;
Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1288 AACATTTTCTTAAGATTTTCAAAAGGAATATGTGTAATTCAGAAATCATAACCACTG 1347
Db 1017 AAAGACAATTTCTTATGATTTGTCATAATTTTAAATCAAAATTTTAGAAATG 1076

QY 1348 TCCTAACTTGGTAAACAAACTGTTCTTAATAAAGTATTTAATGATTTTAAATAAAAAA 1407
Db 1077 AATAAATTCATTCACAAAAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1136

QY 1408 AAAAAAATAAAAAAATAAAAAAATAAAAAA 1434
Db 1137 AAAAAAATAAAAAAATAAAAAAATAAAAAA 1163
```

```
RESULT 10
US-08-461-809-9
; Sequence 9, Application US/08461809
; Patent No. 5770202
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,809
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
```

```
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; LENGTH: 1172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..738
; US-08-461-809-9

Query Match          3.9%; Score 55.8; DB 1; Length 1172;
Best Local Similarity 61.2%; Pred. No. 0.0088;
Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1288 AACATTTTCTTAAGATTTTCAAAAGGAATATGTGTAATTCAGAAATCATAACCACTG 1347
Db 1017 AAAGACAATTTCTTATGATTTGTCATAATTTTAAATCAAAATTTTAGAAATG 1076

QY 1348 TCCTAACTTGGTAAACAAACTGTTCTTAATAAAGTATTTAATGATTTTAAATAAAAAA 1407
Db 1077 AATAAATTCATTCACAAAAATTAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1136

QY 1408 AAAAAAATAAAAAAATAAAAAAATAAAAAA 1434
Db 1137 AAAAAAATAAAAAAATAAAAAAATAAAAAA 1163

RESULT 11
US-08-461-441-9
; Sequence 9, Application US/08461441
; Patent No. 5773002
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,441
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
```

```
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1172 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..738
;   US-08-461-441-9
;
; Query Match          3.9%; Score 55.8; DB 1; Length 1172;
; Best Local Similarity 61.2%; Pred. No. 0.0088;
; Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
;
; QY 1288 AAACATTTTCTAAGATTTTCAAAAGGAATATGTGTAATTCGAAATCATTAACCACTG 1347
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1017 AAAGACAATTTCTTATATGATGTCACATAATTTTAAATCATAAATTTTGAATAATG 1076
;
; QY 1348 TCCTAACTTGGTAAACAACTGTTCTTAATAAAGTATTTTAATGATTTTAAAAA 1407
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1077 AATAAATTCATTACAAAAATTTAAAAA 1136
;
; QY 1408 AAAAAA 1434
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1137 AAAAAA 1163
;
; RESULT 13
; US-09-621-976-16038
; Sequence 16038, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16038
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 281
; OTHER INFORMATION: n=a, g, c or t
; US-09-621-976-16038
;
; Query Match          3.9%; Score 55.6; DB 4; Length 335;
; Best Local Similarity 63.2%; Pred. No. 0.0077;
; Matches 79; Conservative 3; Mismatches 43; Indels 0; Gaps 0;
;
; QY 1310 AAAAGGAATATGTGTAATTCGAAATTCATAACCACTGCTCTAAGTGTAAACAACTG 1369
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 197 AATAGAAAAATAATTCGATAGAAAAATAATAGAAAAATTTTAAAAAACAACCA 256
;
; QY 1370 TTCTTAATAAAGTATTTAATGATTTTAAAAA 1429
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 257 AGCCTATTCTATGAAAAAAMYNVMAA 316
;
; QY 1430 AAAAA 1434
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 317 AAAAA 321
;
; RESULT 14
; US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 593827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
```

```
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1172 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..738
;   US-08-461-441-9
;
; Query Match          3.9%; Score 55.8; DB 1; Length 1172;
; Best Local Similarity 61.2%; Pred. No. 0.0088;
; Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
;
; QY 1288 AAACATTTTCTAAGATTTTCAAAAGGAATATGTGTAATTCGAAATCATTAACCACTG 1347
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1017 AAAGACAATTTCTTATATGATGTCACATAATTTTAAATCATAAATTTTGAATAATG 1076
;
; QY 1348 TCCTAACTTGGTAAACAACTGTTCTTAATAAAGTATTTTAATGATTTTAAAAA 1407
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1077 AATAAATTCATTACAAAAATTTAAAAA 1136
;
; QY 1408 AAAAAA 1434
;      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
; Db 1137 AAAAAA 1163
;
; RESULT 12
; PCT-US93-08518-9
; Sequence 9, Application PC/TUS9308518
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08518
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1172 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..738
```

```
;
;
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Query Match 3.9%; Score 55.4; DB 2; Length 19124;
Best Local Similarity 48.3%; Pred. No. 0.017;
Matches 185; Conservative 0; Mismatches 196; Indels 2; Gaps 1;

QY 1052 CAATGTTAAATTTTTTTTTCATGGAGGCTTTAAATTTGTCTCTTTTCATATTTT 1111
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 17977 CTATTTTCGATTTTTCATTTTTCAGTAAATTTATTTATTTATTTTGTGATTTT 18036
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 11112 ATTCATATTCATTTTATGGTTGTACTGCTTTTGGGAGATAATTAATGTTATAAAT 1171
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 18037 ATAATATATTTTAAATGTGTTATATATATGTTTATATATGTTTATGTTTATTTTGT 18096
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1172 TAGTTTTGGGGGAATAATTTGCAAGAGGATAATTTAATTTACGTCTTCTGTTATTC 1231
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 18097 TACTCTAATCTGAATAAT--CCGAGCGAATAAATAATTAATCTCATATAAATAATTA 18154
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1232 AGAATAAGAGAGAGACTACGCTGCATATTCAGAGTTGTACCTTAACATTTGGTGAAC 1291
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 18155 TTATAATACATATTTATATAGTTTCTTATTAATAAATAATTAATAATACATAATA 18214
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1292 ATTTTCTTAAGATTTTCAAAAGGATATGTGTAATTCGAGAAATCATACCACTGCTCT 1351
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 18215 TTCTTTGTTATTTTATAAATAAATAATTTCTTATTTTATTAACCTTTATTCCTTTT 18274
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1352 AACTTCGTAACCAACTGTTCTTAAATAAGTATTTAATGATTTTAAATAAATAAATAA 1411
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 18275 TAATTTCTTAATTTCTTTATCAAAACAAATAAAGTAATTTCTACATATCAACAAAA 18334
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1412 AAAAAAAAAAAAAAAAAAAAAA 1434
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 18335 AAAAAAAAAAAAAAAAAAAAAA 18357
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-09-621-976-10193
; Sequence 10193, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
```

```
;
;
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10193
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-621-976-10193

Query Match 3.8%; Score 55; DB 4; Length 271;
Best Local Similarity 81.0%; Pred. No. 0.0098;
Matches 64; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1356 TGGTAAACAACTGTTCTTAAATAAAAGTATTTAATGATTTTAAAAA 1415
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1416 AAAAAAAAAAAAAAAAAAAAAA 1434
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 251 AAAAAAAAAAAAAAAAAAAAAA 269
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: July 20, 2004, 18:53:15
Job time : 131 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 18:51:04 ; Search time 676 Seconds
(without alignments)

10349.037 Million cell updates/sec

Title: US-10-799-747-12

Perfect score: 1434

Sequence: 1 catataactcttttattcg.....aaaaaaaaaaaaaaaaaaaaa 1434

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 3191023 seqs, 2439312756 residues

Total number of hits satisfying chosen parameters: 6382046

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1432	99.9	1434	15	US-10-195-730-12
2	425.8	29.7	448	9	US-09-880-107-1385
3	73.2	5.1	560	17	US-10-021-323-2253
4	71.2	5.0	547	17	US-10-021-323-16890
5	70.8	4.9	520	17	US-10-021-323-7699
6	69.6	4.9	517	17	US-10-021-323-11054
7	68.4	4.8	560	17	US-10-021-323-2253
8	67.8	4.7	419	9	US-09-960-352-11234
9	63.6	4.4	528	17	US-10-021-323-8131
10	63.2	4.4	520	17	US-10-021-323-7699
11	62.4	4.4	513	17	US-10-021-323-10547
12	62.4	4.4	537	17	US-10-021-323-7228
13	61.8	4.3	556	17	US-10-021-323-11082
14	61.6	4.3	1284	13	US-10-424-599-25449

c 15	61.2	4.3	393	9	US-09-960-352-4582	Sequence 4582, Ap
c 16	61.2	4.3	508	10	US-09-814-353-5025	Sequence 5025, Ap
c 17	61.2	4.3	508	10	US-09-814-353-11317	Sequence 11317, A
c 18	61.2	4.3	756	10	US-09-814-353-6226	Sequence 6226, Ap
c 19	61.2	4.3	756	10	US-09-814-353-12504	Sequence 12504, A
c 20	61	4.3	469	17	US-10-021-323-16830	Sequence 16830, A
c 21	61	4.3	500	10	US-09-814-353-5831	Sequence 5831, Ap
c 22	61	4.3	500	10	US-09-814-353-12112	Sequence 12112, A
c 23	61	4.3	510	10	US-09-814-353-6245	Sequence 6245, Ap
c 24	61	4.3	510	10	US-09-814-353-12523	Sequence 12523, A
c 25	61	4.3	515	13	US-10-424-599-106126	Sequence 106126, A
c 26	61	4.3	517	17	US-10-021-323-11054	Sequence 11054, A
c 27	60.4	4.2	329	9	US-09-880-107-706	Sequence 706, Ap
c 28	60.2	4.2	517	17	US-10-021-323-16116	Sequence 16116, A
c 29	60.2	4.2	528	17	US-10-021-323-1809	Sequence 1809, Ap
c 30	60.2	4.2	6076	13	US-10-221-714A-386	Sequence 386, App
c 31	60.2	4.2	3673778	15	US-10-312-841-1	Sequence 1, Appli
c 32	60	4.2	6195	15	US-10-311-455-563	Sequence 563, App
c 33	60	4.2	17389	15	US-10-311-455-1388	Sequence 1388, Ap
c 34	59.8	4.2	476	10	US-09-814-353-17420	Sequence 17420, A
c 35	59.8	4.2	6065	13	US-10-221-613-222	Sequence 222, App
c 36	59.4	4.1	756	10	US-09-814-353-6226	Sequence 6226, Ap
c 37	59.4	4.1	756	10	US-09-814-353-12504	Sequence 12504, A
c 38	59.2	4.1	337	9	US-09-960-352-6976	Sequence 6976, Ap
c 39	59.2	4.1	17294	15	US-10-311-455-959	Sequence 959, App
c 40	59	4.1	476	10	US-09-918-995-15561	Sequence 15561, A
c 41	59	4.1	7352	15	US-10-311-455-343	Sequence 343, App
c 42	58.8	4.1	420	10	US-09-814-353-12772	Sequence 12772, A
c 43	58.8	4.1	421	10	US-09-814-353-5169	Sequence 5169, Ap
c 44	58.8	4.1	502	17	US-10-021-323-11431	Sequence 11431, A
c 45	58.8	4.1	10365	15	US-10-098-939-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-10-195-730-12
; Sequence 12, Application US/10195730
; Publication No. US20030144492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: PZ017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-195-730-12

Query Match 99.9%; Score 1432; DB 15; Length 1434;
Best Local Similarity 100.0%; Pred. No. 7.2e-272;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CATTAACTCTTTTATCGGAATAGTAGTATTTTCAATGTCACCTCCATTTCATGTTCA	60
Db	1	CATTAACTCTTTTATCGGAATAGTAGTATTTTCAATGTCACCTCCATTTCATGTTCA	60
QY	61	TTTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTATTTTATATTGAAAGTCAGTGC	120
Db	61	TTTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTATTTTATATTGAAAGTCAGTGC	120
QY	121	AAATATTGAATAGGATATACTAATAATAACAAAGTAAACAAAGTCAGTGT	180

121 AAAAAATGATAGGATATACCTAATTAATACAAAGTAAACAAAGCTCAAGCAGTGT 180
181 CTAATAAATAATCTGGGTTCTTAAATAATTAATTAATAATTAATCTTGAATAATAGTTTCT 240
181 CTAATAAATAATCTGGGTTCTTAAATAATTAATTAATAATTAATCTTGAATAATAGTTTCT 240
241 TAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAGTGTAGTATCATATAAA 300
241 TAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAGTGTAGTATCATATAAA 300
301 CAAATTTCTATTAATGACGAGTGTGATATACAGAAATTTATCAGGCATTAACCAAGTC 360
301 CAAATTTCTATTAATGACGAGTGTGATATACAGAAATTTATCAGGCATTAACCAAGTC 360
361 TAGGCATATAGAAATGAGCACTCAGAAATGGTTTCAATGATAGTATGATGCTGTGA 420
361 TAGGCATATAGAAATGAGCACTCAGAAATGGTTTCAATGATAGTATGATGCTGTGA 420
421 AGGTAGGGAGCTTATTCAGACATAGATAGTATTTCTTAATGCTGTSTCAATTCGTCG 480
421 AGGTAGGGAGCTTATTCAGACATAGATAGTATTTCTTAATGCTGTSTCAATTCGTCG 480
481 CCTTTGCTACCTGTACTTCCCATATATGCGAGCCCATTCAGTCTTGAGTTTCTTCTCT 540
481 CCTTTGCTACCTGTACTTCCCATATATGCGAGCCCATTCAGTCTTGAGTTTCTTCTCT 540
541 GGCACCTTATGCTGAAATCATAGCAGGAGGCTGATTAATTTGGTGAATTTGGGTAGAAA 600
541 GGCACCTTATGCTGAAATCATAGCAGGAGGCTGATTAATTTGGTGAATTTGGGTAGAAA 600
601 GCAGTAGTTTCTGCACATTAAGATCTAGGTATAGATAGTATGAGTATGAGTATGAGTAT 660
601 GCAGTAGTTTCTGCACATTAAGATCTAGGTATAGATAGTATGAGTATGAGTATGAGTAT 660
661 GTTTTTATCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCACCTCTCCCTCCCGCAGC 720
661 GTTTTTATCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCACCTCTCCCTCCCGCAGC 720
721 CTATCTCAAACTGGTGGAAATATATGAGAGATCTTGAAGAGATCTTGAAGAGATTAATAAC 780
721 CTATCTCAAACTGGTGGAAATATATGAGAGATCTTGAAGAGATCTTGAAGAGATTAATAAC 780
781 GCTCCACTCCAGGTGAATCGCCCACTCCCACTGACCTAGTATGATTAATTTAATATAC 840
781 GCTCCACTCCAGGTGAATCGCCCACTCCCACTGACCTAGTATGATTAATTTAATATAC 840
841 TTACCTTCTATTTCTGAAATCAGTGTGAACTGTGCTTATGCTTCAGAGTTTAAAGAAC 900
841 TTACCTTCTATTTCTGAAATCAGTGTGAACTGTGCTTATGCTTCAGAGTTTAAAGAAC 900
901 CTGAGTGAATTCATTTTAAATCTGCTATTCAGAGCCATTAATGATTAATTTCTTAACA 960
901 CTGAGTGAATTCATTTTAAATCTGCTATTCAGAGCCATTAATGATTAATTTCTTAACA 960
961 AGAAGACTCATCTGAGTGTGCTGACTCCCTATGAGCCCATTAAGGTTCTGCTGCTTA 1020
961 AGAAGACTCATCTGAGTGTGCTGACTCCCTATGAGCCCATTAAGGTTCTGCTGCTTA 1020
1021 GCATTAACAAAATAAGGTTATAGGTAAGCCCAATGATTAATTTTGTGATGAGG 1080
1021 GCATTAACAAAATAAGGTTATAGGTAAGCCCAATGATTAATTTTGTGATGAGG 1080
1081 GCTTTAAATTTGCTCTTTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1140
1081 GCTTTAAATTTGCTCTTTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1140
1141 CTTTTAGGAGATAATATATGTTATTAATAATAGTTTTGGGGGGAATAATTTGCAAGA 1200
1141 CTTTTAGGAGATAATATATGTTATTAATAATAGTTTTGGGGGGAATAATTTGCAAGA 1200
1201 GGATTAATTTAATGCTTCTGTTTATTCAGAAATTAAGAGAGAGACTACGCTGCATA 1260

1201 GGATTAATTTAATGCTTCTGTTTATTCAGAAATTAAGAGAGAGACTACGCTGCATA 1260
1261 TTCAAGAGTTGTACCTTAACATTTGGTCAACATTTTCTTAAGATTTTCAAAAGGAATAT 1320
1261 TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATAT 1320
1321 GTGTAATTCAGAAATCATACCACTGTCTTAACCTGGTAAACAACTGTTCTTAATAAA 1380
1321 GTGTAATTCAGAAATCATACCACTGTCTTAACCTGGTAAACAACTGTTCTTAATAAA 1380
1381 AGTATTTAATGATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1434
1381 AGTATTTAATGATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1434

RESULT 2
US-09-880-107-1385/c
; Sequence 1385, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1385
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA491000
US-09-880-107-1385

Query Match 29.7%; Score 425.8; DB 9; Length 448;
Best Local Similarity 98.2%; Pred. No. 4.7e-74;
Matches 441; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

965 GACTCATCTGTAGCTGTTTGTGACTCCTATGAGCCCCCATAGGGTTCTGTGCTTAGCAT 1024
448 GACTCATCTGTAGCTGTTTGTGACTCCTATGAG-CCCATAGGGTTCTGTGCTTAGCAT 390
1025 TAACAAATAAGGTTTATAGGTAAGCCCAATGATTAATTTTGTGATGAGGGCTT 1084
389 TAACAAATAAGGTTTATAGGTAAGCCCAATGATTAATTTTGTGATGAGGGCTT 330
1085 TAAATTTGCTGCTCTTTTTCATATTTTATTCATTTCAATTTATGTTTGTAACTGCTTT 1144
329 TAAATTTGCTGCTCTTTTTCATATTTTATTCATTTCAATTTATGTTTGTAACTGCTTT 270
1145 TTAGGAGATAATATATATGTTTAAATAGTTTGGGGGAATAATTTGCAAAAGAGGA 1204
269 TTAGGAGATAATATATATGTTTAAATAGTTTGGGGGAATAATTTGCAAAAGAGGA 210
1205 AATTTAATTTAGCTGCTTCTGTTTATTCAGAAATTAAGAGAGAGACTACGCTGCATATCA 1264
209 AATTTAATTTAGCTGCTTCTGTTTATTCAGAAATTAAGAGAGAGACTACGCTGCATATCA 150
1265 AGAGTTGTACCTTAACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATATGTTGT 1324
149 AGAGTTGTACCTTAACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATATGTTGT 90
1325 AAATTCAGAAATCATTAACCACTGTCTTAACCTGGTAAACAACTGTTCTTAATAAGTA 1384
89 AAATTCAGAAATCATTAACCACTGTCTTAACCTGGTAAACAACTGTTCTTAATAAGTA 30


```
Query Match 4.9%; Score 70.8; DB 17; Length 520;
Best Local Similarity 49.2%; Pred. No. 0.00036;
Matches 186; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 1057 TATTAAATTTTTCATGAGGCGCTTAAATTTTGTGCTCTTTTCAATTTTATTC 1116
DB 450 TAAAAATATATATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTC 391

QY 1117 TATTCATTTATGTTTGTACGCTTTTTCAGGAGATAATTAATTTAGTTTATTA 1176
DB 390 AAAAAAATATATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTC 331

QY 1177 TTGGGGGAATAATGTGCAAGAGGATAATTAATTTAGCTCTTGTATTCAGAA 1236
DB 330 TTTTAAATTAATTAATTAATTTTTCATCATACATTTTTCATTTTTCATTC 271

QY 1237 AAAGAGAGAGACTACGCTGCATATCAAGAGTTGATACCTTAACATTTGGTCA 1296
DB 270 AAAATACCTTTTTCATCATACATTTTTCATTTTTCATTTTTCATTTTTC 211

QY 1297 TTCTACATTTTCAAGAGGATATGCTTAATTCAGAAATCAACACATTCCTA 1356
DB 210 TTTAAATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTTTTCAT 151

QY 1357 GGTAAACAACTGTTCTTAAATAAGTATTTAATGATTTTAAATTTTAAAT 1416
DB 150 AAAAAAATTAATTAATTAATTTTTCATCATACATTTTTCATTTTTCATTC 91

QY 1417 AAAAAAATTAATTAATTTTTCATCATACATTTTTCATTTTTCATTTTTC 1434
DB 90 AAAAAAATTAATTAATTTTTCATCATACATTTTTCATTTTTCATTTTTCAT 73

RESULT 6
US-10-021-323-11054
; Sequence 11054, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; PRIOR FILING DATE: 2001-12-12
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 11054
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(517)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3828-033-Q6-N6-E7
US-10-021-323-11054

Query Match 4.9%; Score 69.6; DB 17; Length 517;
Best Local Similarity 48.9%; Pred. No. 0.00065;
Matches 186; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 1055 TGATTAATTTTTCATGAGGCGCTTAAATTTTGTGCTCTTTTCAATTTTAT 1114
DB 59 TTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCAT 118

QY 1115 CATATTCAATTTTGTCTACTGCTTTTTCAGGAGATAATTAATTTAGTAA 1174
DB 119 TTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCAT 178
```

```
QY 1175 TTTTGGGGGAATAATTTGTCGAAGAGGATAATTTAATTTACGTGCTTCTGTTATTCAGA 1234
DB 179 TTTTAAAAAATAATTTTAAATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCAT 238

QY 1235 ATAAAGAGAGAGACTACGCTGCATATTTCAAGAGTTGCTTACCTTAACATTTGGTGAACAT 1294
DB 239 AAAAAAATTAATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTTTTCATTC 298

QY 1295 TTTTCTAAGATTTTCAAAGGAATATGTTAAATTTGAGAAATCATACCACTGCTCTTAAC 1354
DB 299 AAAAAAATTAATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTTTTCATTC 358

QY 1355 TTGTTAAACAACTGTTCTTAATTAATTAATTAATTTTTCATTTTTCATTTTTCATTTTTC 1414
DB 359 TTTAAATTAATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTTTTCATTTTTC 418

QY 1415 AAAAAAATTAATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTTTTCAT 1434
DB 419 TAAAAAAGGAAAAAATTAATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCAT 438

RESULT 7
US-10-021-323-2253/c
; Sequence 2253, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 2253
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3825-011-Q1-N6-C9
US-10-021-323-2253

Query Match 4.8%; Score 68.4; DB 17; Length 560;
Best Local Similarity 48.9%; Pred. No. 0.0012;
Matches 183; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 1061 AATTTTTCATGAGGCGCTTAAATTTTGTGCTCTTTTTCATTTTTCATTTTTCATTT 1120
DB 374 ATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTC 315

QY 1121 CAATTTATGTTTGTAACTGCTTTTTCAGGAGATAATTAATTTAGTTTTCGTTG 1180
DB 314 TTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCAT 255

QY 1181 GGGGAATAATTTGCAAGAGGATAATTTAATTTACGTGCTTCTGTTATTCAGAAATAAG 1240
DB 254 TAATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCATTTTTC 195

QY 1241 AGAAGACATACGCTGCATATTTCAAGAGTTGATACCTTAACATTTGGTGAACATTTTCT 1300
DB 194 ATAAATTAATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTTTTCATTTTTC 135

QY 1301 AAGATTTTCAAAGGAATATGTTAAATTTGAGAAATCATACCACTGCTCTTAACCTGGTA 1360
DB 134 AAAAAAATTAATTTTTCATCATACATTTTTCATTTTTCATTTTTCATTTTTCATTTTTC 75

QY 1361 AACAACTGTTCTTAATTAAGTATTTAATGATTTTTCATTTTTCATTTTTCATTTTTCAT 1420
```

Db 74 AA 15
Qy 1421 AAAAAAAAAAAAAAAAAA 1434
Db 14 AAAAAAAAAA 1

RESULT 8

US-09-960-352-11234/c
; Sequence 11234, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11234
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-LIB3058-052-Q1-K1-D8
US-09-960-352-11234

Query Match 4.7%; Score 67.8; DB 9; Length 419;
Best Local Similarity 51.0%; Pred. No. 0.0013;
Matches 187; Conservative 0; Mismatches 177; Indels 3; Gaps 1;
Qy 1068 TTTTGCATGGAGGCTTTAAATTTGGCTCTTTTCATATTTTATTCATATTCATATTTA 1127
Db 415 TTTTAAATATATATTTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 356
Qy 1128 TGGTTGTAACTGCTTTTGGGAGATAATATATATGTATATAATTTAGTTTGGGGGAAT 1187
Db 355 ATATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTA 296
Qy 1188 AATTGCGAAGAGGATAATTTTAACTGCTCTGTTTATTCAGATAATAAGAGAGAAG 1247
Db 295 TTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAA 239
Qy 1248 ACTACGCTGATATTCAGAGTTGTACCTTAACTTGGTGAAACATTTTCTAAGATTT 1307
Db 238 TTTTATATATTTTAAATAATTTTAAATAATTTTAAATAATTTTAAATAATTTTAA 179
Qy 1308 TCAAAAGGAATATGTGTAATTCAGAAATCATAACCACTGCTCTGTAACCAAC 1367
Db 178 TTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 119
Qy 1368 TGTTCTTAATAAGTATTTTAAATTTTAAATAATAATAATAATAATAATAATAA 1427
Db 118 AA 59
Qy 1428 AAAAAAA 1434
Db 58 AAAAAA 52

RESULT 9

US-10-021-323-8131/c
; Sequence 8131, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 8131
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3828-006-Q1-N6-F8
US-10-021-323-8131

Query Match 4.4%; Score 63.6; DB 17; Length 528;
Best Local Similarity 48.1%; Pred. No. 0.01;
Matches 171; Conservative 0; Mismatches 179; Indels 0; Gaps 0;
Qy 1082 CTTTAAATTTGGCTCTTTTCATATTTTATTCATATTCATATTCATATTCATATTC 1141
Db 350 CCTAAATAATTTTTTTTTTTTTTTTTTAAACCCCTTTTTTTTTTTTTTTTTT 291
Qy 1142 TTTTAGGAGATAAT 1201
Db 290 TTTTATTAAACCCCTTTTT 231
Qy 1202 GATAATTTAATTTACGTGCTCTCTGTTTATTCAGATAATAAGAGAGAGACTACGTG 1261
Db 230 TTTTAAATAATAATTTTTTTTTTTTTTTTTTTTTTAAATAATAATAATAATAATAA 171
Qy 1262 TCAGAGTTGTACCTTAACTTAACTTGGTGAAACATTTTCTAAGATTTTCAAGAGATATG 1321
Db 170 TTTTATTTTTTTTTTAAATAATAATAATAATAATAATAATAATAATAATAATAA 111
Qy 1322 TGTAATTTGAGAAATCATAACCACTGCTCTGTAACCTGTTTCTTAAATATAA 1381
Db 110 AA 51
Qy 1382 GTATTTAATGATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAA 1431
Db 50 AA 1

RESULT 10

US-10-021-323-7699
; Sequence 7699, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 7699
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3828-019-Q1-N6-D6
US-10-021-323-7699

Query Match 4.4%; Score 63.2; DB 17; Length 520;
Best Local Similarity 48.1%; Pred. No. 0.012;
Matches 179; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
Qy 1063 TTTTATTTTGCATGGAGGCTTTAAATAATTTGTGCTCTTTTCATATTTTATTCATATCA 1122

[illegible]

RESULT 14
 US-10-424-599-25449
 ; Sequence 25449, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 25449
 ; LENGTH: 1284
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_122983C.1
 ; US-10-424-599-25449

Query Match	4.3%	Score 61.6;	DB 13;	Length 1284;
Best Local Similarity	50.7%;	Pred. No. 0.038;		
Matches 148;	Conservative 0;	Mismatches 144;	Indels 0;	Gaps 0;
QY	1143	TTTTAGGGAGATAATTATGTTATAAATTAGTTTTCGGGGGGAATAATTGTGCAAGAGG	1202	
Db	640	TGTTAGAGAAAAGGATTATTTCGTGCTTTCCTTGTTCGAGAGCAAAAAATTGTGAAAAAGGT	699	
QY	1203	ATAATTTTAATTTACGTGCTTCTGTTATTTCAGATAAAGAGAGAGAAGCTACGCTGCATATT	1262	
Db	700	ATTGAAGATCCAGAAGGCAAAAGCAAGCAGGCATCAAGAGTTAAGCTGTTGAGTTATT	759	
QY	1263	CAAGAGTTGTACTTAAACATTTGTGTGAACACATTTTTCTAAGATTTTCAAAAGGAATATGT	1322	
Db	760	CGGGAGTGAACAAAAAGATCTCTGTTAAAAATTTGTGCTAAAGAAATTTTGATTTTGAATGGTG	819	
QY	1323	GTAAATTGAGAAATCATAAACCACCTGCTCTAACTTGGTAAACAAACACTGTTCTTAAATAAAG	1382	

```

Db      820   TTAAGTATGAATGTGAAACCTTAATTGAAATTATTTATGAAAAAATAAAAAAAAAAAAAA 879
Qy      1383   TATTTAATGATTTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
          |||
Db      880   AAAAAAAGGAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 931
          |||

RESULT 15
US-09-960-352-4582/c
; Sequence 4582, Application US/09960352
; Patent NO. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Machialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATE
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4582
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 20-LIB3057-013-Q1-K1-E7
US-09-960-352-4582

```

Query Match	4.3%;	Score 61.2;	DB 9;	Length 393;
Best Local Similarity	52.8%;	Fred. No. 0.026;		
Matches 132;	Conservative 0;	Mismatches 118;	Indels 0;	Gaps 0;
QY	1185	AATAATTGTGCAAGAGGATAATTAAATTTACGTGCTTCTGTTATTCAGAAATAAGAGAG	1244	
DB	366	AAAAATTAAATATAAAAAAABAAATTTATAAAAAATATTTAAATTTATAATAATAAAATAAT	307	
QY	1245	AAGCTACGCTGCATATTCAGAGTGTGCTACCTTCAACATTCGGTGAACATTTTTTCTTAAGA	1304	
DB	306	ATAATCATAAAATATAATAATATATTTTAAATAATTTTTTAAAAAATATATATTTAAAAA	247	
QY	1305	TTTTTCAAAGGAATATGTGTAATTTGAGAAATCATAAACCACGTGCTTAACCTTGGTAAACA	1364	
DB	246	AAAAAABAAATTTTATAAATAATTTTTTAAAAATAAAAAAATATATAAATAAAAAA	187	
QY	1365	AACGTGCTTCAAAATAAGATTTATTAATCGATTTTAAABAAAAAABAAAAAABAAAAA	1424	
DB	186	TAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAAABAAAAA	127	
QY	1425	AAAAAABAAAA	1434	
DB	126	AAAAAABAAAA	117	

Search completed: July 20, 2004, 21:22:37
Job time : 680 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 17:01:24 ; Search time 3467 Seconds
(without alignments)
12351.418 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaaactcttttcatcg.....aaaaaaaaaaaaaaaaaaaa 1434

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	871.2	60.8	1201	13 BX441923	BX441923 BX441923
2	744.6	51.9	814	14 CB853053	CB853053 UI-CF-FNO
3	643.6	44.9	765	12 B1914473	B1914473 603182264
4	606	42.3	785	12 BG484396	BG484396 602505037

C	5	576.2	40.2	598	10	BF431622	BF431622 7016e08.x
C	6	551.4	38.5	568	9	AL035942	AL035942 DRFP564B
	7	549.8	38.3	575	12	BM722991	BM722991 UI-E-E01-
	8	543.6	37.9	557	14	CB132625	CB132625 K-EST0183
C	9	537.2	37.5	556	10	AW500190	AW500190 UI-HF-BNO
	10	534	37.2	551	12	BT259796	BT259796 602971440
	11	513.8	35.8	621	9	AL035941	AL035941 DRFP564B
	12	468.2	32.6	503	10	AW889139	AW889139 QV4-NT002
C	13	444	31.0	444	9	AA136080	AA136080 zK90D03.s
C	14	427	29.8	438	10	AW304923	AW304923 xV97b07.x
C	15	425.8	29.7	448	9	AA491000	AA491000 aa529g11.s
C	16	415.2	29.0	1037	13	EX413464	EX413464 BX413464
C	17	396.4	27.6	400	9	AW002842	AW002842 wt56h11.x
C	18	395	27.5	395	9	AA992811	AA992811 ot85D03.s
C	19	372.6	26.0	688	12	BM728856	BM728856 UI-E-E01-
C	20	370.6	25.8	394	9	AA747222	AA747222 nx76a10.s
C	21	354.6	24.7	419	10	AW189910	AW189910 x110a04.x
C	22	349.4	24.4	493	9	AW026264	AW026264 wt10h02.x
C	23	323.8	22.6	941	13	EX452125	EX452125 BX452125
C	24	295.4	20.6	957	12	BG403642	BG403642 602419361
C	25	278.4	19.4	280	9	AA679646	AA679646 ah10c06.s
C	26	273	19.0	338	12	BI025209	BI025209 CM4-NT028
C	27	233	16.2	253	9	AA953459	AA953459 on69g09.s
C	28	168.8	11.8	240	10	BF910533	BF910533 CM4-UT004
C	29	88.6	6.2	1084	13	EX361825	EX361825 BX361825
C	30	84.6	5.9	1126	13	EX446388	EX446388 EX446388
C	31	84.4	5.9	1101	29	CNS00EQL	AL069526 Drosophil
C	32	83.6	5.8	1133	13	EX444099	EX444099 BX444099
C	33	82.4	5.7	634	13	BX506904	BX506904 DRFP2779F
C	34	81.6	5.7	1056	13	EX415058	EX415058 BX415058
C	35	81.4	5.7	1101	29	CNS00EVL	AL069706 Drosophil
C	36	81	5.6	1200	13	EX437758	EX437758 BX437758
C	37	80.4	5.6	625	29	CNS036A2	AL229763 Tetradon
C	38	80.4	5.6	1200	13	EX436510	EX436510 BX436510
C	39	79.4	5.5	1043	29	CNS0145P	AL103735 Drosophil
C	40	78.6	5.5	681	13	BX378802	BX378802 BX378802
C	41	77.8	5.4	1101	29	CNS00EVL	AL069706 Drosophil
C	42	77.8	5.4	1201	13	EX402521	EX402521 BX402521
C	43	77.4	5.4	450	29	CNS020QX	AL207042 Tetradon
C	44	76.6	5.3	999	13	BX380865	BX380865 BX380865
C	45	76	5.3	681	13	BX378802	BX378802 BX378802

ALIGNMENTS

RESULT 1
BX441923
LOCUS BX441923 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS0DF023Y002 5-PRIME, mRNA sequence.
ACCESSION BX441923
VERSION BX441923.1 GI:30771989
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)
Li W.B., Gruber C., Jessee J. and Polayes D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2532.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF023BH01QPI&cluster=2532.r. Contact :
Feng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DF023BH01QPI.

FEATURES

source

```
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF023Y002"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
```

ORIGIN

```
Query Match      60.8%; Score 871.2; DB 13; Length 1201;
Best Local Similarity 93.0%; Pred. No. 2.3e-104;
Matches 937; Conservative 21; Mismatches 37; Indels 13; Gaps 4;

QY 415 CTTCTAGGTAGGGAGCTTATTCAGACATAGTAGATGTTCTCTAATGCTGTSTCAAT 474
Db |||||
QY 54 CTTCTAGGTAGGGAGCTTATTCAGACATAGTAGATGTTCTCTAATGCTGTCTCAAT 113
Db |||||
QY 475 TGCTGGCCTTTGGCTACCTGCTACTTCCSCATTATGGCAGCCCAATTCAGTCTTGAGTTTC 534
Db |||||
QY 114 TGCTGGCCTTTGGCTACCTGCTACTTCCSCATTATGGCAGCCCAATTCAGTCTTGAGTTTC 173
QY 535 TTCTCTGACACCTTATGCTCTGAAATCATGAGGAGGCTGATTCATTTGGTGGTGGG 594
Db |||||
QY 174 TTCTCTGACACCTTATGCTCTGAAATCATGAGGAGGCTGATTCATTTGGTGGTGGG 233
QY 595 TAGAAGCAGTATGTTTGTGTCATTAAGTGTAGGTATAGATAGTTTAGCTTTAA 654
Db |||||
QY 234 TAGAAGCAGTATGTTTGTGTCATTAAGTGTAGGTATAGATAGTTTAGCTTTAA 293
QY 655 GTGATGTTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCACCTCTCC 714
Db |||||
QY 294 GTGATGTTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCACCTCTCC 353
QY 715 CCCAGCCTATCTCAAACTGGTGGATATATGAGAGATCTTGAAGAAGTAAATAAAC 774
Db |||||
QY 354 CCCAGCCTATCTCAAACTGGTGGATATATGAGAGATCTTGAAGAAGTAAATAAAC 413
QY 775 TTCACTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGATTTGTAAT 834
Db |||||
QY 414 TTCACTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGATTTGTAAT 473
QY 835 TAATACTTACTTCTATTTCTGAATCAGTTGTGAACCTGTTGCCTTATGTTTCAGAGTTT 894
Db |||||
QY 474 TAATACTTACTTCTATTTCTGAATCAGTTGTGAACCTGTTGCCTTATGTTTCAGAGTTT 533
QY 895 AAGAACCTGCTGAATTCATTTTAAATCTGCTATTCGAGAGCAATTCGAATGAATTC 954
Db |||||
QY 534 AAGAACCTGCTGAATTCATTTTAAATCTGCTATTCGAGAGCAATTCGAATGAATTC 593
QY 955 TTAACAAGAGACTCATCTGCTAGCTGTTTGTGACTCTATGAGCCCAATAGGGTCTG 1014
Db |||||
QY 594 TTAACAAGAGACTCATCTGCTAGCTGTTTGTGACTCTATGAGCCCAATAGGGTCTG 652
QY 1015 TGCTTAGCATTAACAAATAAGTTTATAGTTAAAGCCCAATGATTTAA-TTTTTTTTTTC 1073
Db |||||
QY 653 TGCTTAGCATTAACAAATAAGTTTATAGTTAAAGCCCAATGATTTAA-TTTTTTTTTTC 712
QY 1074 ATGAGGGCTTTAAATTTGCTGCTTTTTCATATTTTATCATATTCATTTATGTTT 1133
Db |||||
QY 713 ATGAGGGCTTTAAATTTGCTGCTTTTTCATATTTTATCATATTCATTTATGTTT 772
QY 1134 GTAACCTGTTTTTAGGGAGATAATATATGTTTAAATAGTTTGGGGGGAATAATTGT 1193
Db |||||
QY 773 GTAACCTGTTTTTAGGGAGATAATATATGTTTAAATAGTTTGGGGGGAATAATTGT 832
Db |||||
```

```
QY 1194 GCBAAGAGGATAATTTAATTTACGTGCTTCTGTTATTTCAGATAAAGAGAGACTACG 1253
Db |||||
QY 833 GCAAGAGGATAATTTAATTTACGTGCTTCTGTTATTTCAGATAAAGAGAGACTACG 892
Db |||||
QY 1254 CTGCATATTTCAAGAGTGTCTACCTTAACATTCGTGTAACATTTTCTTAAGATTTTCAAAA 1313
Db |||||
QY 893 CTGCATATTTCAAGAGTGTCTACCTTAACATTCGTGTAACATTTTCTTAAGATTTTCAAAA 951
Db |||||
QY 1314 GGAATATGTTAAATTTGAGAAATCATACCACTGCTCTAATCTTGGTAAACAACTGTTCT 1373
Db |||||
QY 952 GAAATGCTGTTTTTAAATTTGAGAAATCATACCACTGCTCTAATCTTGGTAAACAACTGTTCT 1001
QY 1374 TAAATAAGATTTTAAATGATTTTAAAAAATAAAAAAATAAAAAAATAAAAAA 1421
Db |||||
QY 1002 CTAATAAGATTTTAAATGATTTTAAATGATTTTAAAAAATAAAAAAATAAAAAAATAAAAAA 1049
Db |||||
```

RESULT 2

```
CB853053/c
LOCUS CB853053 814 bp mRNA linear EST 22-APR-2003
DEFINITION UI-CF-FNO-afg-e-20-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone
ACCESSION CB853053
VERSION CB853053.1 GI:30048171
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 814)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Bernaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com)
The following repetitive elements were found in this cDNA
sequence: 1-42, >AT-rich#low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.
```

FEATURES

```
Location/Qualifiers
source 1. 814
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-FNO-afg-e-20-0-UI"
/tissue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies)" (T1 phage resistant)"
/clone_lib="UI-CF-FNO"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-FNO is a subtracted cDNA library derived from two
normalized Human lung epithelial cell libraries (EN1 and
DUI) The library was subtracted according to according to
Bernaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. For additional information, contact:
bento-soares@uiowa.edu
TAG_SEQ=None found"
```

ORIGIN

Query Match 51.9%; Score 744.6; DB 14; Length 814;
 Best Local Similarity 98.9%; Pred. No. 8.2e-88;
 Matches 796; Conservative 2; Mismatches 2; Indels 5; Gaps 5;

QY	610	TTTGTCTGACATTAAAGATGTAGGTTATAGATAGGTTTGTAGCCCTTTAAAGTGTATGTTTATA	669
DB	800	TTTGTCTGACATTAAAGATGTA-GTTATAGATA-GTTTGTAGCCTTTAAGTGTATG-TTTTATA	744
QY	670	CTTTTAAATTAAGAAATATAACCTTTTAAAGCTATTCCACCTTCCTCCCGACGCTATCTCAA	729
DB	743	CTTTTAAATTAAGAAATATAACC-TTTTAAAGCTATTCCACCTTCCTCCCGACGCTATCTCAA	685
QY	730	ACTGGTGGAAATATATCGAGAGATCTTGTGAAGAGTAGTAAATAAAACCTTTCACCTGCTCCACCTC	789
DB	684	ACTGGTGGAAATATATCGAGAGATCTTGTGAAGAGAGTAAATAAAACCTTTCACCTGCTCCACCTC	625
QY	790	CAGGTGAATCCGCCCACTCCCACTGACCTAGTAGAGAAATTTGTAAATTTAAATCTTTACCTTCT	849
DB	624	CAGGTGAATCCGCCCACTCCCACTGACCTAGTAGAGAAATTTGTAAATTTAAATCTTTACCTTCT	565
QY	850	ATTTCTGAATACAGTTGTGAACGTGTGCTTATGTTTCAGARGTTTAAAGACCTCMGTGAA	909
DB	564	ATTTCTGAATACAGTTGTGAACGTGTGCTTATGTTTCAGARGTTTAAAGACCTCMGTGAA	505
QY	910	TTCAATTTTTTAAATCTGCTATTCTCGAGAAGCATTTGAATGAAATCTTTTAAACAAGAAGCTC	969
DB	504	TTCAATTTTTTAAATCTGCTATTCTCGAGAAGCATTTGAATGAAATCTTTTAAACAAGAAGCTC	445
QY	970	ATCTGTAGCTGTTTGTGTAUCCCTATGAGCCCAATAAGGGTTCTGTGCTTTAGCATTTAAACA	1029
DB	444	ATCTGTAGCTGTTTGTGTAUCCCTATGAGCCCAATAAGGGTTCTGTGCTTTAGCATTTAAACA	386
QY	1030	AAATAAGGTTTATAGGTAAGCCAAATGATTAATTTTTTTTTTTCGATCGAGGGCTTTAAAA	1089
DB	385	AAATAAGGTTTATAGGTAAGCCAAATGATTAATTTTTTTTTTTCGATCGAGGGCTTTAAAA	326
QY	1090	TTTGTGCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTGTGAACCTGCTTTTATG	1149
DB	325	TTTGTGCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTGTGAACCTGCTTTTATG	266
QY	1150	GAGATAATTTATATGTTTATAAATTTAGTTTTCGGGGGAAATAATGTGCAAGAGGATAATTT	1209
DB	265	GAGATAATTTATATGTTTATAAATTTAGTTTTCGGGGGAAATAATGTGCAAGAGGATAATTT	206
QY	1210	AATTTAGTCTCTCTGTTTATTCAGATAAAGAGAGAGACTACGCTCGCATATTCAGAGT	1269
DB	205	AATTTAGTCTCTCTGTTTATTCAGATAAAGAGAGAGACTACGCTCGCATATTCAGAGT	146
QY	1270	TGTACCTTACATTTGGTGAACAATTTTCTTAAGATTTTCAAAAGGAATATGTGTAATTT	1329
DB	145	TGTACCTTACATTTGGTGAACAATTTTCTTAAGATTTTCAAAAGGAATATGTGTAATTT	86
QY	1330	GAGAAATCATACCACTGCTCCTAACCTTGGTAAACAAACCTGTTCTTAATTAAGATTTTAA	1389
DB	85	GAGAAATCATACCACTGCTCCTAACCTTGGTAAACAAACCTGTTCTTAATTAAGATTTTAA	26
QY	1390	TGATTTTTTAAAAAATTT	1414
DB	25	TGATTTTTTAAAAAATTT	1

RESULT 3
 BI914473
 LOCUS
 DEFINITION BI914473 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246463 5',
 mRNA sequence.
 ACCESSION BI914473
 VERSION BI914473.1
 KEYWORDS GI:16178652
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 765)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: sgabbs@mail.nih.gov
Tissue procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11621 row: p column: 16
High quality sequence stop: 762.
Location/Qualifiers
1..765
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5246463"
/lab_host="DH10B"
/clone_lib="NIH_MGC_121"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 017. Note:
this is a NIH MGC Library."

Query Match	44.9%;	Score 643.6;	DB 12;	Length 765;
Best Local Similarity	97.8%;	Pred. No. 1.1e-74;		
Matches 702;	Conservative 3;	Mismatches 7;	Indels 6;	Gaps 5;
QY	1	CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTCACATCCATTCAATGTTGA	60	
DB	49	CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTCACATCCATTCAATGTTGA	107	
QY	61	TTTCGAGCTGACACGCTTATTTTGTCTAAAGCAGACAGATTTAAATTTTATATTGAAAGTCAGTGC	120	
DB	108	TTTCGAGCTGACACGCTTATTTTGTCTAAAGCAGACAGATTTAAATTTTATATTGAAAGTCAGTGC	167	
QY	121	AAAATTTAGTATAGGATATACATAAATAACAAAGTAAATACAAAAGTCAAAAGCAGTCTT	180	
DB	168	AAAATTTAGTATAGGATATACATAAATAACAAAGTAAATACAAAAGTCAAAAGCAGTCTT	227	
QY	181	CTAATTAATAAATCTCGGGTTCCTTAAATAATTTTAAATTTATC-TTGAATAATGTTTTC	239	
DB	228	CTAATTAATAAATCTCGGGTTCCTTAAATAATTTTAAATTTATCTTTGAATAATGTTTTC	287	
QY	240	TTAGATTAATCTCAGGATATCAGAAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA	299	
DB	288	TTAGATTAATCTCAGGATATCAGAAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA	347	
QY	300	ACAAATTTGCTATTAAATGTCAMGA-GTGGTAAATACAGAAATTTATCAGGCATTACCAAG	358	
DB	348	ACAAATTTGCTATTAAATGCAAGACGTGGTAAATACAGAAATTTATCAGGCATTACCAAG	407	
QY	359	TCTAGGCACATATAGGAAATGCAGCACTCAGAAATGGTTCAATCTAGTATGTTGATGCTTG	418	
DB	408	TCTAGGCACATATAGGAAATGCAGCACTCAGAAATGGTTCAATCTAGTATGTTGATGCTTG	467	
QY	419	TAAAGTAGGGAGCTTATTTCAGACATAGTATAGTATGTTCTCTAATGCTGTSTCAATTGCT	478	
DB	468	TAAAGTAGGGAGCTTATTTCAGACATAGTATAGTATGTTCTCTAATGCTGTSTCAATTGCT	527	
QY	479	GGCCTTTTGGCTACCTGTACTTCCSCATTATGGACCCCAATTCAGTCTTGAGTTTCTTCT	538	

Db 528 GGCCCTTGGCTACCTGTACTTCCCATTTATGSCAGCCCATTCAGTCTTGAGTTTCTCT 587
 QY 539 CTGGACACCTTATGCTCTGAATCATGAGCGAGGCTGATTCATTAATTTGGGTAGG 598
 Db 588 CTGGACACCTTATGCTCTGAATCATGAGCGAGGCTGATTCATTAATTTGGGTAGG 646
 QY 599 AAGCAGTAGTGTCTGACATTAAGATGTAGGTATAGATAGATAGATAGATAGATAG 658
 Db 647 AAGCAGTAGTGTCTGACATTAAGATGTAGGTATAGATAGATAGATAGATAGATAG 706
 QY 659 AAGTGTGTATATCTTAAATTAAGAAATATAACCTTTTAAGCTATTCACCTCTCCCC 716
 Db 707 ATG--TCTATACTTTAAATAAGAAATATAACCTTTTAAGCTATTCACCTCTCCCC 762

RESULT 4
 BG484396 785 bp mRNA linear EST 21-MAR-2001
 LOCUS 6025037F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618473 5',
 DEFINITION mRNA sequence.

ACCESSION BG484396
 VERSION BG484396.1 GI:13416675

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 785)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM1376 row: j column: 10

High quality sequence start: 18

High quality sequence stop: 751.

Location/Qualifiers

1. 785

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4618473"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH_MGC_77"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:

Sf11 (ggccgctcgcc); Site:2: Sf11 (ggccattatggcc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CACGGCATATGCCC-3' and 3' adaptor sequence:

5'-ATTGAGGCGGCGGCGCATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.9

kb (range 0.5-4.0 kb). 12/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech laboratories (Palo Alto,

CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 42.3%; Score 606; DB 12; Length 785;
 Best Local Similarity 95.4%; Pred. No. 7.9e-70;
 Matches 682; Conservative 5; Mismatches 21; Indels 7; Gaps 6;

QY 249 TCTCAGGATGAGAAAGTCAATTAAGTGTAGTAAAGTTAGTATCATTAACAAATGT 308

Db 63 TCTAGAGATGACACCGTCCAGTAAGTGTGAGTAAAGTTAGTATCATTAACAAATGT 122

QY 309 CTATTAATGCMGA-GTGGTAATATACAGATTTATCAGGCATTACCAAGTCTAGGCAC 367

Db 123 CTATTAATGCMGACGCTGGTAATATACAGATTTATCAGGCATTACCAAGTCTAGGCAC 182
 QY 368 ATATAGGAAATGAGCACTCAGAATGTTTCAATGTAGTGTGATGCTGTGTAAGGTAGG 427
 Db 183 ATATAGGAAATGAGCACTCAGAATGTTTCAATGTAGTGTGATGCTGTGTAAGGTAGG 242
 QY 428 GGAGCTTATTCAGACATAGTAGATTTTCTTAATGCTGTSTCAATTTGCTGGCCTTTGG 487
 Db 243 GGAGCTTATTCAGACATAGTAGATTTTCTTAATGCTGTCTCAATTTGCTGGCCTTTGG 302
 QY 488 CTACTGTACTTCCSCATTATGGCAGCCCATTCAGTCTTGTAGTGTCTTCTCTCTGACACC 547
 Db 303 CTACTGTACTTCCCATTTATGGCAGCCCATTCAGTCTTGTAGTGTCTTCTCTGACACC 362
 QY 548 TTATGCTCTGAAATCATGAGCGAGCTGATTCATTTGGTGTGATTTGGGTAGAAAGCAGTAT 607
 Db 363 TTATGCTCTGAAATCATGAGCGAGCTGATTCATTTGGTGTGATTTGGGTAGAAAGCAGTAT 422
 QY 608 GTTTTGTCTGACATTAAGATGTAGGTATAGATAGTATTAGCTTTAAGTGTATGTTTTA 667
 Db 423 GTTTTGTCTGACATTAAGATGTAGGTATAGATAGTATTAGCTTTAAGTGTATGTTTTA 482
 QY 668 TACTTTAAATAAGAAATATATAACCTTTTAAGCTATTCACCTCTCTCCCGCAGCTATCTC 727
 Db 483 TACTTTAAATAAGAAATATATAACCTTTTAAGCTATTCACCTCTCTCCCGCAGCTATCTC 542
 QY 728 AAACCTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCACTGCTCCAC 787
 Db 543 AAACCTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCACTGCTCCAC 602
 QY 788 TCCAGGTGAATCCCGCCACTCCCACTGACCTAGTAGTAA-TTGTGAATTTAATCTTACT 846
 Db 603 TCCAGGTGAATCCCGCCA-TCCCACTGACCTAGTAGAATTTTGAATTTAATCTTACT 661
 QY 847 TCTATTTCTGAAATCAGTTGTGAAT-CTTGCCCTTATGTTTCAGAGTGTTAAGAACCCTCMG 905
 Db 662 TCTA-TTCTGAAATCAGTTGTGAATCTTGCTTATGTTTCAGAGTGTTAAGAACCCTCAG 720
 QY 906 TGAATTCATTTTAAATCTGCTA-TTCTGAGAAGCATTGAATGAAATCTTTAA 958
 Db 721 TGAATTCATTTTAAATCTGCTAATTTCTGAGGAGGCAATGACTGAATCTTTA 775

RESULT 5

BF431622/c

LOCUS

7016e08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3574359 3',

DEFINITION mRNA sequence.

ACCESSION BF431622

VERSION BF431622.1 GI:11443736

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 588)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Sequencing by: Greg Lemmon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

infoimage.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 469.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 568)
 Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
 EST (Wambutt, et al.)
 Unpublished (1999)
 Contact: MIPS
 MIPS
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 3' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 xl sequence also available.
 This clone (DKFzp564B1622) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers
 1. 568
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFzp564B1622"
 /tissue_type="brain"
 /dev_stage="fetal"
 /lab_host="Xl-2blue"
 /clone_l1b="564 (synonym: hibr2)"
 /note="Vector: pAMPI; Site_1: NotI; Site_2: SalI"
 ORIGIN
 Query Match 38.5%; Score 551.4; DB 9; Length 568;
 Best Local Similarity 98.9%; Pred. No. 1.2e-62;
 Matches 563; Conservative 2; Mismatches 3; Indels 1; Gaps 1;
 QY 844 CCTTCTATTTCGAAATCAGTGTGAACTGTTGCTTATGTTTCAGAGGTTTAAAGACCTC 903
 DB 568 CCTTCTATTTCGAAATCAGTGTGAACTGTTGCTTATGTTTCAGAGGTTTAAAGACCTC 509
 QY 904 MGTGAATTCATTTTAAAAATCTGCTATTCGAGAGCATGTAATGAATTCCTTAAACAAGA 963
 DB 508 AGTGAATTCATTTTAAAAATCTGCTATTCGAGAGCATGTAATGAATTCCTTAAACAAGA 449
 QY 964 AGACTCATCTGAGCTGTTGCTGACTCCCTATGAGCCCAATAGGGTCTGTGCTTAGCA 1023
 DB 448 AGACTCATCTGAGCTGTTGCTGACTCCCTATGAG-CCCAATAGGGTCTGTGCTTAGCA 390
 QY 1024 TTAACAAAATAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTTCATGAGAGGCT 1083
 DB 389 TTAACAAAATAAGTATTAGGTAAGCCCAATGTAATTAATTTTTCATGAGAGGCT 330
 QY 1084 TTAATAATTGTCCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTGTAACTGCTT 1143
 DB 329 TTAATAATTGTCCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTGTAACTGCTT 270
 QY 1144 TTTTAGGAGATATATATGTTATAATAGTTTTCGGGGGAATATTTGTCGAAGAGGA 1203
 DB 269 TTTTAGGAGATATATATGTTATAATAGTTTTCGGGGGAATATTTGTCGAAGAGGA 210
 QY 1204 TAATTTAATTTACGTGCTTCTGTTATTTCAAGATAAGAGAGAGACTACGCTGCATATTC 1263
 DB 209 TAATTTAATTTACGGCTCTCTGTTATTCAGAAATAAGAGAGAGACTACGCTGCATATTC 150
 QY 1264 AAGAGTTGTACCTTAACATTTGGTGAACATTTTTTCTAAGATTTTCAAAAGGAATATGTG 1323
 DB 149 AAGAGTTGTACCTTAACATTTGGTGAACATTTTTTCTAAGATTTTCAAAAGGAATATGTG 90
 QY 1324 TAAATTGAGAAATCATAAACACTGCTCTAACTTGGTGAACAAACTGTTCTTAATAAAGT 1383
 DB 89 TAAATTGAGAAATTAATAACCACTGCTCTAACTTGGTGAACGAACTGTTCTTAATAAAGT 30

```

QY 1384 ATTGAATGATTTTAAAAAAGAAAAA 1412
Db 29 ATTGAATGATTTTAAAAAAGAAAAA 1
|||||
BM722991 575 bp mRNA linear EST 01-MAR-2002
UI-E-EOI-aid-a-14-0-UI.r1 UI-E-EOI Homo sapiens cDNA clone
UI-E-EOI-aid-a-14-0-UI 5', mRNA sequence.
BM722991
ACCESSION BM722991
VERSION 1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 575)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 148-169, >AT-rich#Low_complexity (matched complement)
Seq primer: M13 Reverse.
Location/Qualifiers
1..575
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EOI-aid-a-14-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EOI"
/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EOI is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCGTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."
```

Query Match 38.3%; Score 549.8; DB 12; Length 575;
Best Local Similarity 99.1%; Pred. No. 1.9e-62;
Matches 570; Conservative 3; Mismatches 0; Indels 2; Gaps 2;

QY 56 GTTGATTTGGAGCTGACAGTATTTTGTGTAAGCAGAGATTTAATTTATATTGAAGTC 115

```

Db 1 GTTGATTTGGAGCTGACAGTATTTTGTGTAAGCAGAGATTTAATTTATATTGAAGTC 60
|||||
QY 116 AGTGCAAAATATGAATAGGATATATAATAAATACAAAGTAATAACAAAAGTCAAGCA 175
|||||
Db 61 AGTGCAAAATATGAATAGGATATATAATAAATACAAAGTAATAACAAAAGTCAAGCA 120
|||||
QY 176 GTGTTCTAAATAAATAATCTCGGGTTCCTTAAATAATTTTAAATTTATC-TTGAATAAG 234
|||||
Db 121 GTGTTCTAAATAAATAATCTCGGGTTCCTTAAATAATTTTAAATTTATC-TTGAATAAG 180
|||||
QY 235 TTTTCTTAGATTAAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGCTTAGTATC 294
|||||
Db 181 TTTTCTTAGATTAAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGCTTAGTATC 240
|||||
QY 295 ATTAACAAATTTGTTCTATTAAATGCMGA-GTGTGTAATATACAGATTTTATCAGGCATTA 353
|||||
Db 241 ATTAACAAATTTGTTCTATTAAATGCAAGCGTGTGTAATATACAGATTTTATCAGGCATTA 300
|||||
QY 354 CCAAGTCTAGGCACATATAGGAATGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 413
|||||
Db 301 CCAAGTCTAGGCACATATAGGAATGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 360
|||||
QY 414 GCTTGTAAAGTAGGGAGCTTATTTCAGACATAGTAGATAGTTTCTCTAATGCTGTSTCAA 473
|||||
Db 361 GCTTGTAAAGTAGGGAGCTTATTTCAGACATAGTAGATAGTTTCTCTAATGCTGTSTCAA 420
|||||
QY 474 TTGCTGGCCTTTGGCTACCTGCTACCTGCTACCTGCTACCTGCTACCTGCTACCTGCTACCTG 533
|||||
Db 421 TTGCTGGCCTTTGGCTACCTGCTACCTGCTACCTGCTACCTGCTACCTGCTACCTGCTACCTG 480
|||||
QY 534 CTCTCTGAGCACCTTATGCTCTGAATCATGAGGAGGCTGATTCATTTGTTGATTTGG 593
|||||
Db 481 CTCTCTGAGCACCTTATGCTCTGAATCATGAGGAGGCTGATTCATTTGTTGATTTGG 540
|||||
QY 594 GTAGAAAGCAGTATGTTTGTGCTGACATTAAGATGT 628
|||||
Db 541 GTAGAAAGCAGTATGTTTGTGCTGACATTAAGATGT 575
|||||
```

RESULT 8
CB132625
LOCUS CB132625
DEFINITION K-EST0183064 L14ChoiCK0 Homo sapiens cDNA clone L14ChoiCK0-2-D11
5', mRNA sequence.
ACCESSION CB132625
VERSION CB132625.1 GI:28097898
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 557)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 2 row: D column: 11
High quality sequence stop: 557.
Location/Qualifiers
1..557
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L14ChoiCK0-2-D11"

TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 2 row: D column: 11
High quality sequence stop: 557.
Location/Qualifiers
1..557
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L14ChoiCK0-2-D11"

FEATURES
source
1..557
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L14ChoiCK0-2-D11"

/sex="M"
/cell_line="Choi-Ck"
/lab_host="Top10F"
/clone_lib="L14ChoiCK0"
/notes="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;
Site 2: NotI; The library was contributed by the Soares
Laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN

Query Match 37.9%; Score 543.6; DB 14; Length 557;
Best Local Similarity 99.5%; Pred. No. 1.3e-61;
Matches 554; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
Source 1.556

674 AATAAGAAATATACCTTTTAAAGCTATTCCACCTCTCCGCCAGCCTATCTCAACTG 733
Db 1 AATAAGAAATATACCTTTTAAAGCTATTCCACCTCTCCGCCAGCCTATCTCAACTG 60
734 GTGGAATATATGAGAGATCTTGAAGAAGTAAATAAACTTCACCTGCTCCACTCCAGG 793
Db 61 GTGGAATATATGAGAGATCTTGAAGAAGTAAATAAACTTCACCTGCTCCACTCCAGG 120
794 TGAATCGGCCACCTCCCACTGACCTAGTAGAATTTGTAATTAATTAATTAATTAATTT 853
Db 121 TGAATCGGCCACCTCCCACTGACCTAGTAGAATTTGTAATTAATTAATTAATTAATTT 180
854 CTGAATACAGTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTT 913
Db 181 CTGAATACAGTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTT 240
914 TTTTAAAAATCTGCTATTCTGAGAAGCAATGAATGAATTTTAAACAAGAGACTCATCT 973
Db 241 TTTTAAAAATCTGCTATTCTGAGAAGCAATGAATGAATTTTAAACAAGAGACTCATCT 300
974 GTAGCTGTTTGTGACCTCTATGAGCCCATTAAGGGTTCTGTGCTTAGCAATTAACAAAT 1033
Db 301 GTAGCTGTTTGTGACCTCTATGAG-CCCATTAAGGGTTCTGTGCTTAGCAATTAACAAAT 359
1034 AAGTTTATAGTTAAGGCAATGATTAATTTTGTGCTAGAGAGGCTTTAAATTTG 1093
Db 360 AAGTTTATAGTTAAGGCAATGATTAATTTTGTGCTAGAGAGGCTTTAAATTTG 419
1094 TGCTCTTTTCAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1153
Db 420 TGCTCTTTTCAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 479
1154 TAATTTATGTTATAAATTTAGTTTGGGGGAATTAATTTGCAAGAGATTAATTTAATTT 1213
Db 480 TAATTTATGTTATAAATTTAGTTTGGGGGAATTAATTTGCAAGAGATTAATTTAATTT 539
1214 TACGTGCTCTGTTATTC 1231
Db 540 TACGTGCTCTGTTATTC 557

RESULT 9
AW500190/c
LOCUS
DEFINITION UI-HF-BNO-akc-g-07-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
INAME:3076789 5', mRNA sequence.
ACCESSION AW500190
VERSION AW500190.1 GI:7112570
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 556)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
The following repetitive elements were found in this cDNA
sequence:
8-48, >AT-rich#Low complexity
Seq primer: M13 Forward.
Location/Qualifiers
1.556

FEATURES

Source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3076789"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="WGC85"
/lab_host="DH10B (LTI)"
/clone_lib="NIH_MGC_50"
/notes="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(3.5-4.4kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

ORIGIN

Query Match 37.5%; Score 537.2; DB 10; Length 556;
Best Local Similarity 99.5%; Pred. No. 8.5e-61;
Matches 547; Conservative 2; Mismatches 0; Indels 1; Gaps 1;
Source 1.556

864 TTGTGAACCTGCTGCTTATGTTTCTGAGAGTTTAAAGACCTCMGTGAATTCATTTTAAAA 923
Db 556 TTGTGAACCTGCTGCTTATGTTTCTGAGAGTTTAAAGACCTCMGTGAATTCATTTTAAAA 497
924 TCTGCTATTCTGAGAGCAATGAATGAATTTCTTAAACAGAGAGCTCATCTGTAGCTGTTT 983
Db 496 TCTGCTATTCTGAGAGCAATGAATGAATTTCTTAAACAGAGAGCTCATCTGTAGCTGTTT 437
984 GCTGACTCCTATGAGCCCCATAAGGGTTCTGTGCTTAGCATTAAACAAAATAAGGTTTATA 1043
Db 436 GCTGACTCCTATGAG-CCCATTAAGGGTTCTGTGCTTAGCATTAAACAAAATAAGGTTTATA 378
1044 GGTAAAGCCCAATGATTAATTTTCTGATGAGAGGCTTTAAATTTGCTCTCTTTT 1103
Db 377 GGTAAAGCCCAATGATTAATTTTCTGATGAGAGGCTTTAAATTTGCTCTCTTTT 318
1104 CATATTTTATTCATATTCAATTTTATGTTTGTAACTCTTTTAGGGAGATAATATATG 1163
Db 317 CATATTTTATTCATATTCAATTTTATGTTTGTAACTCTTTTAGGGAGATAATATATG 258
1164 TTATAAATTAGTTTGGGGGAATAATTTGTGCAAGAGGATAATTTAATTTAGCTGCTTC 1223
Db 257 TTATAAATTAGTTTGGGGGAATAATTTGTGCAAGAGGATAATTTAATTTAGCTGCTTC 198
1224 TGTATTTCAGAAATAAGAGAGAGAGACATACCTGCTGATTTCAAGAGTTGTACCTTAACATT 1293
Db 197 TGTATTTCAGAAATAAGAGAGAGAGACATACCTGCTGATTTCAAGAGTTGTACCTTAACATT 138
1284 GGTGAACATTTTCTTAAGATTTTCAAAAGGATATGTTGAAATTCAGAAATCATAAACC 1343
Db 137 GGTGAACATTTTCTTAAGATTTTCAAAAGGATATGTTGAAATTCAGAAATCATAAACC 78
1344 ACTGTCCTAACTTGGTAAACAAACTGTTCTTAAATAAGATTTAATGATTTTAAAAAAA 1403
Db 77 ACTGTCCTAACTTGGTAAACAAACTGTTCTTAAATAAGATTTAATGATTTTAAAAAAA 18
1404 AAAAAAAAAA 1413

```

Db      17 AAAAAAAAAA 8
|||||
BI259796      551 bp mRNA linear EST 17-JUL-2001
602971440Fl NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5110904 5',
mRNA sequence.
BI259796
BI259796.1 GI:148117472
EUT.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 551)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11268 row: p column: 09
High quality sequence stop: 536.
FEATURES             source
Location/Qualifiers
1..551
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5110904"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_lib="NIH_MGC_12"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
ORIGIN
Query Match      37.2%; Score 534; DB 12; Length 551;
Best Local Similarity 99.1%; Pred. No. 2.2e-60;
Matches 545; Conservative 2; Mismatches 2; Indels 1; Gaps 1;
Qy      880 TATGTTTCAGAGTTTAAAGAACCTCMGTGAATTCATTTTAAATCTGCTATTCGAGAA 939
Db      1 TATGTTTCAGAGTTTAAAGAACCTCAGTGAATTCATTTTAAATCTGCTATTCGAGAA 60
Qy      940 GCATTGAATGAAATCTTAAAGAGAGACTCATCTGTAGCTGTTTGTGCTGACTTCCTATGAGC 999
Db      61 GCATTGAATGAAATCTTAAAGAGAGACTCATCTGTAGCTGTTTGTGCTGACTTCCTATGAG- 119
Qy      1000 CCCATAGGGTCTGCTGCTTAGCATTAAACAAATAAGTTTATAGGTTAAAGCCAAATGAT 1059
Db      120 CCCATAGGGTCTGCTGCTTAGCATTAAACAAATAAGTTTATAGGTTAAAGCCAAATGAT 179
Qy      1060 TAAATTTTTTTTGCATGGAGGGCTTTAAATAATTTGCTCTTTTTCATATTTTATTCATAT 1119
Db      180 TAAATTTTTTTTGCATGGAGGGCTTTAAATAATTTGCTCTTTTTCATATTTTATTCATAT 239
Qy      1120 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTAATGTTATAAATTAAGTTTGT 1179
Db      240 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTAATGTTATAAATTAAGTTTGT 299
Qy      1180 GGGGGAATAATTTGCAAGAGGAGTAATTTAAATTTAGTCTGCTCTGTTATTCAGAAATAA 1239
Db      300 GGGGGAATAATTTGCAAGAGGAGTAATTTAAATTTAGTCTGCTCTGTTATTCAGAAATAA 359

Qy      1240 GAGAGAGACTACGCTGCATATTTCAAGAGTTGTACCTTAAACATTCGTTGTAACATTTTTC 1299
Db      360 GAGAGAGACTACGCTGCATATTTCAAGAGTTGTACCTTAAACATTCGTTGTAACATTTTTC 419
Qy      1300 TAAAGATTTTCAAAAAGGAATATGTGTAAATTCAGAAATCATACCACTGTCTTAACTTGT 1359
Db      420 TAAAGATTTTCAAAAAGGAATATGTGTAAATTCAGAAATCATACCACTGTCTTAACTTGT 479
Qy      1360 AAACAACACTGTTCTTTAAATAAAGTATTTTAAATGATTTTAAAAAATAAAAAAATAAAA 1419
Db      480 AAACAACACTGTTCTTTAAATAAAGTATTTTAAATGATTTTAAAAAATAAAAAAATAAAA 539
Qy      1420 AAAAAAATAAAA 1429
Db      540 AAAAAAATAAAA 549

RESULT 11
LOCUS      AL035941      621 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp564B1622 r1 564 (synonym: hEbr2) Homo sapiens cDNA clone
ACCESSION AL035941
VERSION AL035941.1 GI:5405575
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 621)
Wambutt, R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
EST (Wambutt, et al.)
Unpublished (1999)
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
s1 sequence also available.
This clone (DKFZp564B1622) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             source
Location/Qualifiers
1..621
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp564B1622"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="X1-2blue"
/clone_lib="564 (synonym: hEbr2)"
/note="Vector: pAMPI; Site 1: NotI; Site 2: SalI"
ORIGIN
Query Match      35.8%; Score 513.8; DB 9; Length 621;
Best Local Similarity 99.1%; Pred. No. 8.5e-58;
Matches 534; Conservative 3; Mismatches 0; Indels 2; Gaps 2;
Qy      1 CATTAACTCTTTTATCGGGAATAGTATGATATTTCAATGTCACCTCCATTCATGTTGA 60
Db      83 CATTAACTCTTTTATCGGGAATAGTATGATATTTCAATGTCACCTCCATTCATGTTGA 142
Qy      61 TTTGGAGCTGACAGTATTTTGTGTAGCAGAGATTTAAATTTTATATTGAAGTCAGTGC 120
Db      143 TTTGGAGCTGACAGTATTTTGTGTAGCAGAGATTTAAATTTTATATTGAAGTCAGTGC 202
Qy      121 AAAATTATGAATAGGATATCTACTAATAATAACAAGTAAATAACAAAGTCACAAAGCAGTGT 180

```

Db 203 AAAATTATGATAGGATATCTAATAATAAATAAAGTAATAAAGTCAAAGTCAAGTGT 262

QY 181 CTAAATAAATAATCTGGTCTCTTTAAATAATTTATTTAAATTTATC-TTGAATAAGTTTTTC 239

Db 263 CTAAATAAATAATCTGGTCTCTTTAAATAATTTATTTAAATTTATC-TTGAATAAGTTTTTC 322

QY 240 TTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATTA 299

Db 323 TTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATTA 382

QY 300 ACAAATGTCTATTAATGACAG- GTGGTAATATATACAGAAATTTATCAGGCATACCAAG 358

Db 383 ACAAATGTCTATTAATGACAGCGTGGTAATATATACAGAAATTTATCAGGCATATACCAAG 442

QY 359 TCTAGGCACATATAGGAAATGACGACCTCAGATGAGTTTCAATGTAGTGTGATGCTTG 418

Db 443 TCTAGGCACATATAGGAAATGACGACCTCAGATGAGTTTCAATGTAGTGTGATGCTTG 502

QY 419 TAAAGTAGGGAGCTTATTCAGACATAGTATAGTATTTCTCTAATGCTGTGTCATTTGCT 478

Db 503 TAAAGTAGGGAGCTTATTCAGACATAGTATAGTATTTCTCTAATGCTGTGTCATTTGCT 562

QY 479 GGCTTTGGCTACTGCTTCTCCATATGGCAGCCCATTCAGTCTTGAGTTTCTTCTC 537

Db 563 GGCTTTGGCTACTGCTTCTCCATATGGCAGCCCATTCAGTCTTGAGTTTCTTCTC 621

RESULT 12

AW889139 503 bp mRNA linear EST 24-MAY-2000

LOCUS QV4-NT0028-160400-173-b05 NT0028 Homo sapiens cDNA, mRNA sequence.

DEFINITION AW889139

ACCESSION AW889139.1 GI:8053344

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 503)

AUTHORS Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.P., Matsukuma, A., Baia, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,

O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LiCR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=4t2-QV4-NT0028-160

400-173-b05&t3=2000-04-16&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 18

High quality sequence stop: 502.

Location/Qualifiers

1. .503

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="NT0028"

/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 32.6%; Score 468.2; DB 10; Length 503;
Best Local Similarity 98.0%; Pred. No. 8.1e-52;
Matches 493; Conservative 2; Mismatches 5; Indels 3; Gaps 2;

QY 549 TATGCTCTGAATCATGAGGAGGCTGATTCATTTGGTGTATTTGGGTAGAAAGCAGTATG 608

Db 4 TTTGCTCTGAATCATGAGGAGGCTGAT--CAITGGTATTTGGGTAGAAAGCAGTATG 61

QY 609 TTTTGTGTCATTAAGATGTAGCTTATAGATAGTGTTAGCTTTAAAGTGTATGTTTTTAT 668

Db 62 TTTTGTGTCATTAAGATGTAGCTTATAGATAGTGTTAGCTTTAAAGTGTATGTTTTTAT 121

QY 669 ACTTTAAATAAAGAAATATAACCTTTTAAAGCTATTCACCTCTCCGCCAGCCTATCTCA 728

Db 122 ACTTTAAATAAAGAAATATAACCTTTTAAAGCTATTCACCTCTCCGCCAGCCTATCTCA 181

QY 729 AACTGTGGAATATATGGAGAGATCTTGAAGAGTAAATAAATTAACCTTCAGCTGCTCCAT 788

Db 182 AACTGTGGAATATATGGAGAGATCTTGAAGAGTAAATAAATTAACCTTCAGCTGCTCCAT 241

QY 789 CCAGGTGAATCCGCCACCTCCCACTGACCTAGTAGAATTTGTAATTTAACTTACCTTC 848

Db 242 CCAGGTGAATCCGCCACCTCCCACTGACCTAGTAGAATTTGTAATTTAACTTACCTTC 301

QY 849 TATTTCTGAATCAGTTGTGAACCTGTTGCCCTTATGTTTCAGAGTTTAAAGAACCTCAGTGA 908

Db 302 TATTTCTGAATCAGTTGTGAACCTGTTGCCCTTATGTTTCAGAGTTTAAAGAACCTCAGTGA 361

QY 909 ATTCAATTTTAAATCTGCTATCTGAGAGCATTAATGAATGAATTTTACACAGACACT 968

Db 362 ATTCAATTTTAAATCTGCTATCTGAGAGCATTAATGAATGAATTTTACACAGACACT 421

QY 969 CATCTGTAGCTGTTTGTGCTGACTCTCTATGAGCCCATAGGGTCTCTGTGCTTAGCATTAAC 1028

Db 422 CATCTGTAGCTGTTTGTGCTGACTCTCTATGAG-CCCATAGGGTCTCTGTGCTTAGCATTAAC 480

QY 1029 AAAATAGGTTTATAGGTTAAAGC 1051

Db 481 AAAATAGGTTTATAGGTTAAAGC 503

RESULT 13

AA136080/c

LOCUS

DEFINITION

zk90d03.s1 Soares pregnant uterus_NbHPU Homo sapiens cDNA clone

IMAGE:490085 3', mRNA sequence.

AA136080

ACCSSION

VERSION

AA136080.1 GI:1697290

KEYWORDS

SOURCE

EST.

ORGANISM

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

AUTHORS

1 (bases 1 to 444)

Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 366.

FEATURES

source

1. 444
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3804672"
/db_xref="taxon:9606"
/clone="IMAGE:490085"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Soares pregnant uterus NhPU"
/note="Organ: uterus; Vector: pT773-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo (dt) primer [5',
AAGTGAAGATTCCGCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT773 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

ORIGIN

Query Match 31.0%; Score 444; DB 9; Length 444;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 963 AAGACTCATCTGTAGCTTTGCTGACCCCTATGAGCCCATAGGCTTCTGTGCTAGC 1022
Db 444 AAGACTCATCTGTAGCTTTGCTGACCCCTATGAGCCCATAGGCTTCTGTGCTAGC 385

QY 1023 ATTAACAAATPAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTTCATCGAGGC 1082
Db 384 ATTAACAAATPAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTTCATCGAGGC 325

QY 1083 TTTAAATTTGCTCTTTTTCATATTTTATCATATTTCAATTTATGTTTGTAACTGCT 1142
Db 324 TTTAAATTTGCTCTTTTTCATATTTTATCATATTTCAATTTATGTTTGTAACTGCT 265

QY 1143 TTTTAGGAGATAATTAATGTTTATAAATAGTTTTCGGGGGAATAATTTGCAAGAGG 1202
Db 264 TTTTAGGAGATAATTAATGTTTATAAATAGTTTTCGGGGGAATAATTTGCAAGAGG 205

QY 1203 ATAAATTTTACGTCCTTCTGTTTATCAGAAATAAGAGAGAGACTACGCTGCATATT 1262
Db 204 ATAAATTTTACGTCCTTCTGTTTATCAGAAATAAGAGAGAGACTACGCTGCATATT 145

QY 1263 CAAGAGTTGTACCTTAACATTTGTGAACATTTTTCAGATTTTCAAAGGAATAGT 1322
Db 144 CAAGAGTTGTACCTTAACATTTGTGAACATTTTTCAGATTTTCAAAGGAATAGT 85

QY 1323 GTAAATTCAGAAATCAATAACACCTGCTTAACTTGGTAAACAACTGTTCTTAATAAAG 1382
Db 84 GTAAATTCAGAAATCAATAACACCTGCTTAACTTGGTAAACAACTGTTCTTAATAAAG 25

QY 1383 TATTAAATGATTTTAAAAA 1406
Db 24 TATTAAATGATTTTAAAAA 1

RESULT 14

AW304923/c 438 bp mRNA linear EST 20-JAN-2000
LOCUS xv97b07.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2826421 3',
DEFINITION mRNA sequence.
ACCESSION AW304923
VERSION AW304923.1 GI:67117126
KEYWORDS EST.

SOURCE

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 438)

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BTGP), Tumor Gene Index

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyt not found

Seq primer: -40UP from Gibco

High quality sequence stop: 410.

Location/Qualifiers

1. 438

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2826421"

/tissue_type="three pooled meningiomas"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Brn53"

/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies."

ORIGIN

Query Match 29.8%; Score 427; DB 10; Length 438;
Best Local Similarity 99.8%; Pred. No. 2e-46;
Matches 438; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 961 AGAAGACTCATCTGTAGCTGTTGCTGACCTCTATGAGCCCATAGGCTTCTGTGCTTA 1020
Db 438 AGAAGACTCATCTGTAGCTGTTGCTGACCTCTATGAG-CCCATAGGCTTCTGTGCTTA 380

QY 1021 GCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTTCATGAGG 1080
Db 379 GCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTTCATGAGG 320

QY 1081 GCCTTAAATTTGCTCTTTTTCATATTTTATCATATTTCAATTTATGTTTGTAACTG 1140
Db 319 GCCTTAAATTTGCTCTTTTTCATATTTTATCATATTTCAATTTATGTTTGTAACTG 260

QY 1141 CTTTTTCAGGAGATAATTAATGTTTATAAATAGTTTTCGGGGGAATAATTTGCAAGA 1200
Db 259 CTTTTTCAGGAGATAATTAATGTTTATAAATAGTTTTCGGGGGAATAATTTGCAAGA 200

QY 1201 GGATAATTTTAAATTTACGTCCTTCTGTTTATTCAGATAAGAGAGAGACTACGCTG 1260
Db 199 GGATAATTTTAAATTTACGTCCTTCTGTTTATTCAGATAAGAGAGAGACTACGCTG 140

QY 1261 TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTTCAGATTTTCAAAGGAATAT 1320
Db 139 TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTTCAGATTTTCAAAGGAATAT 80

QY 1321 GTGTAATTTGAGAAATCAATAACACCTGCTTAACTTGGTAAACAACTGTTCTTAATAA 1380
Db 79 GTGTAATTTGAGAAATCAATAACACCTGCTTAACTTGGTAAACAACTGTTCTTAATAA 20

QY 1381 AGTATTAAATGATTTAAA 1399
Db 19 AGTATTAAATGATTTAAA 1

RESULT 15
AA491000/c
LOCUS
DEFINITION
aa52g11.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824612 3',
EST.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 448)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 436.
Location/Qualifiers
1. .448
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:824612"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GCB1"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGGGGCGGCTCATTTTCTTTTCTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

Search completed: July 20, 2004, 19:51:20
Job time : 3474 secs

FEATURES
source

ORIGIN
Query Match 29.7%; Score 425.8; DB 9; Length 448;
Best Local Similarity 98.2%; Pred. No. 2.8e-46;
Matches 441; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 965 GACTCATCTGTAGCTGTTTGTGCTACTCTATGAGCCCATAGGGTCTGTGCTTAGCAT 1024
DB 448 GACTCATCTGTAGCTGTTTGTGCTACTCTATGAG-CCCATAGGGTCTGTGCTTAGCAT 390
QY 1025 TAACAAAATAGGTTTATAGGTAAAGCCATGATTAATTTTTCATGAGGGGCTT 1084
DB 389 TAACAAAATAGGTTTATAGGTAAAGCCATGATTAATTTTTCATGAGGGGCTT 330
QY 1085 TAAATTTTGTCTCTTTTTCATATTTTATCATATTTCAATTTATGCTTTGTAACGCTTT 1144
DB 329 TAAATTTTGTCTCTTTTTCATATTTTATCATATTTCAATTTATGCTTTGTAACGCTTT 270
QY 1145 TTAGGAGATTAATATGTTTATAAATTAGTTTTGGGGGGAATAATTGTGCAAGAGGAT 1204

DB 269 TTAGGAGATTAATATGTTTATAAATTAGTTTTGGGGGGAATAATTGTGCAAGAGGAT 210
QY 1205 AATTTAAATTTACGTGCTTCTGTATTTCAGAATAAAGAGAGAGAGACTAGCTGCATATTC 1264
DB 209 AATTTAAATTTACGTGCTTCTGTATTTCAGAATAAAGAGAGAGAGACTAGCTGCATATTC 150
QY 1265 AGAGTTGTACCTTAAACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATATGTGT 1324
DB 149 AGAGTTGTACCTTAAACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATATGTGT 90
QY 1325 AAATTTGAGAAATCATAACCACTGTCTTAACCTTGGTAAACAAAACCTGTCTTAAATAAAGTA 1384
DB 89 AAATTTGAGAAATCATAACCACTGTCTTAACCTTGGTAAACAAAACCTGTCTTAAATAAAGTA 30
QY 1385 TTATAATGATTTTAAAAAAGAAAAAAGAAAAA 1413
DB 29 TTTGATGATGTTTAAAAAAGAAAAAAGAAAAA 1

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 19:51:29 ; Search time 5499 Seconds
(without alignments)
11302.766 Million cell updates/sec

Title: US-10-799-747-12

Perfect score: 1434
Sequence: 1 cattaactcttttctcg.....aaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1434	100.0	1434	6	BD078422	BD078422 101 human
2	735	51.3	3522	9	HS0804706	AL833393 Homo sapi
3	700	48.8	175081	9	AC021151	AC021151 Homo sapi
4	695	48.5	2020	9	HS0800227	AL049442 Homo sapi
5	631	44.0	3452	9	BC033650	BC033650 Homo sapi
6	400	27.9	281662	2	AC117374	AC117374 Homo sapi
7	345	24.1	448	6	AX408738	AX408738 Sequence
8	187	13.0	281662	2	AC117374	AC117374 Homo sapi
9	142	9.9	395	6	AX898296	AX898296 Sequence
10	142	9.9	395	6	BD033829	BD033829 Sequence
11	47	3.3	1648	10	BC063167	BC063167 Rattus no
12	47	3.3	8281	9	HS0801490	AL133623 Homo sapi
13	46	3.2	1378	3	AK113242	AK113242 Ciona int
14	46	3.2	110000	2	PF011316	Continuation (17 o
15	45	3.1	1648	9	BC009267	BC009267 Homo sapi
16	45	3.1	1772	10	BC058245	BC058245 Mus muscu
17	45	3.1	1922	9	AF232009	AF232009 Homo sapi
18	45	3.1	2156	9	AK025435	AK025435 Homo sapi
19	45	3.1	2240	9	BC036449	BC036449 Homo sapi
20	45	3.1	7084	9	HS0803370	AL832063 Homo sapi
21	44	3.1	1353	3	AY070501	AY070501 Drosophil
22	44	3.1	1452	3	AY058474	AY058474 Drosophil
23	44	3.1	1730	6	E52152	E52152 Gene encodi
24	44	3.1	1730	6	AR361006	AR361006 Sequence
25	44	3.1	1990	10	BC042668	BC042668 Mus muscu
26	44	3.1	2058	9	HS0802809	AL390184 Homo sapi
27	44	3.1	2515	10	BC006577	BC006577 Mus muscu
28	44	3.1	2539	5	BC045290	BC045290 Danio rer
29	44	3.1	2899	9	BC028198	BC028198 Homo sapi
30	44	3.1	3128	9	HS0807508	BX647364 Homo sapi
31	44	3.1	3487	10	BC016095	BC016095 Mus muscu
32	44	3.1	4075	9	BC021300	BC021300 Homo sapi
33	44	3.1	129517	2	AC016400	AC016400 Homo sapi
34	44	3.1	130503	2	AC130694	AC130694 Mus muscu
35	44	3.1	139756	2	AC144592	AC144592 Medicago
36	44	3.1	157274	2	AC140724	AC140724 Mus muscu
37	44	3.1	195221	10	AL683878	AL683878 Mouse DNA
38	44	3.1	211485	2	AC131304	AC131304 Mus muscu
39	44	3.1	213308	2	AC079164	AC079164 Mus muscu
40	44	3.1	237290	2	AC109543	AC109543 Rattus no
41	43	3.0	222	6	AR122013	AR122013 Sequence
42	43	3.0	384	6	AX185690	AX185690 Sequence
43	43	3.0	399	6	AR085693	AR085693 Sequence
44	43	3.0	399	6	I18358	I18358 Sequence 13
45	43	3.0	399	6	I21345	I21345 Sequence 13

ALIGNMENTS

RESULT 1
LOCUS BD078422 1434 bp DNA linear PAT 27-AUG-2002
DEFINITION 101 human secretory proteins.
ACCESSION BD078422
VERSION BD078422.1 GI:22624025
KEYWORDS JP 2001519156-A/11.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1434)
Duan,R.D., Florence,K.A., Rosen,C.A., Ruben,S.M., Greene,J.M.,
Young,P., Ferrie,A.M., Yu,G.L., Janat,F.Ni,J., Carter,K.C.,
Endress,G.A., Feng,P., Lafleur,D.W. and Shi,Y.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3522)
Ottewaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and Wiemann, S.
Direct Submission
Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MediGenomix (Martinried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp762K109) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.
Location/Qualifiers
source
1. 3522
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp762K109"
/db_xref="taxon:9606"
/clone="DKFZp762K109"
/tissue type="melanoma (MeWo cell line)"
/clone_lib="762 (synonym: hm12). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal
3420..3425
polyA_site
3440
ORIGIN

Query Match 51.3%; Score 735; DB 9; Length 3522;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1105; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

324 GTGGTAAATACAGAAATTTATCAGGCATTACCAAGCTAGGCACATATAGCAATGCGC 383
DB GTGGTAAATACAGAAATTTATCAGGCATTACCAAGCTAGGCACATATAGCAATGCGC 2428

384 ACTCAGAAATGTTCAATGAGTAGTGTGATGCTTTGTAAGGTAGGGAGCTTTATTCAGACA 443
DB ACTCAGAAATGTTCAATGAGTAGTGTGATGCTTTGTAAGGTAGGGAGCTTTATTCAGACA 2488

444 TAGTAGATGTTTCTCAATGCTGTSTCAATGCTGCGCTTTGGCTACCTGTACTTCCSC 503
DB TAGTAGATGTTTCTCAATGCTGTSTCAATGCTGCGCTTTGGCTACCTGTACTTCCSC 2548

504 ATTATGCGACCCATTGAGTCTTGAGTTTCTCTCGACACCTATGCTCTGAAATCA 563
DB ATTATGCGACCCATTGAGTCTTGAGTTTCTCTCGACACCTATGCTCTGAAATCA 2608

564 TGAGCGAGGCTGATTCATTTGGTGAATTTGGGTAGAAAGCAGTAGTGTGCTGACATTAA 623
DB TGAGCGAGGCTGATTCATTTGGTGAATTTGGGTAGAAAGCAGTAGTGTGCTGACATTAA 2668

624 GATGAGGTTATAGATAGGTTAGCCCTTAAGCTATGATGTTTATATCTTTAAATAAGAA 683
DB GATGAGGTTATAGATAGGTTAGCCCTTAAGCTATGATGTTTATATCTTTAAATAAGAA 2728

684 ATATAACCTTTAAGCTATTCCACCTCTCCCGACGCTATCTCAAACTGCTGGAATATA 743
DB ATATAACCTTTAAGCTATTCCACCTCTCCCGACGCTATCTCAAACTGCTGGAATATA 2788

744 TGAGGAGATCTTGAAGAAGTAAATAAACCCTTCACTGCTCCACTCCAGGTGAATCGCC 803
DB TGAGGAGATCTTGAAGAAGTAAATAAACCCTTCACTGCTCCACTCCAGGTGAATCGCC 2848

804 CACTCCCACTGACCTAGTAGAATTTGTAATTAATTAATTAATTAATTAATTAATTAAT 863
DB CACTCCCACTGACCTAGTAGAATTTGTAATTAATTAATTAATTAATTAATTAATTAAT 2908

864 TTGTGAATCTGTGCTTATGTTTACAGAGTTTAAAGAACCTTCMGTAATTCATTTTAAAA 923
DB TTGTGAATCTGTGCTTATGTTTACAGAGTTTAAAGAACCTTCMGTAATTCATTTTAAAA 2968

924 TCTGCTATTCTGAGAGCAATTAATGAATCTTAAACAAGAGACTCATCTGTAGCTGTT 983
DB TCTGCTATTCTGAGAGCAATTAATGAATCTTAAACAAGAGACTCATCTGTAGCTGTT 3028

984 GCTGACTCTTATGAGCCCAATAGGGTCTGTGCTTAGCAATTAACAAAAAAGGTTTATA 1043
DB GCTGACTCTTATGAG-CCCAATAAGGGTCTGTGCTTAGCAATTAACAAAAAAGGTTTATA 3087

1044 GGTAAAGCCCAATGTAATTAATTTTTCATCGAGGGCTTTAAAAATTTGTCCTTTT 1103
DB GGTAAAGCCCAATGTAATTAATTTTTCATCGAGGGCTTTAAAAATTTGTCCTTTT 3147

1104 CATATTTTATTCATATTCAATTTATGTTTGAATCTTTTAGGAGAGATAATATATG 1163
DB CATATTTTATTCATATTCAATTTATGTTTGAATCTTTTAGGAGAGATAATATATG 3207

1164 TTATAAATTAGTTTGGGGGGAATAATTGTGCAAGAGGATAATTTAATTTAGTGCTTC 1223
DB TTATAAATTAGTTTGGGGGGAATAATTGTGCAAGAGGATAATTTAATTTAGTGCTTC 3267

1224 TGTATTTCAGAAATAAGAGAGAGAGACTACGCTGCATATTCAGAGTGTACCTTAACATT 1283
DB TGTATTTCAGAAATAAGAGAGAGAGACTACGCTGCATATTCAGAGTGTACCTTAACATT 3327

1284 GGTGAACCAATTTTCTTAAGATTTTCAAAAGAGAAATATGTAAATTCAGAAATCATAACC 1343
DB GGTGAACCAATTTTCTTAAGATTTTCAAAAGAGAAATATGTAAATTCAGAAATCATAACC 3387

1344 ACTGCTCTTAATGTTGTAACAAACTGTTCTTAAATAAGAGTGTAAATTTAATGATTTAAAAA 1403
DB ACTGCTCTTAATGTTGTAACAAACTGTTCTTAAATAAGAGTGTAAATTTAATGATTTAAAAA 3447

1404 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
DB AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3478

RESULT 3
AC021151/c 175081 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-483A20 from 4, complete sequence.
DEFINITION AC021151
ACCESSION AC021151.8 GI:15145598
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175081)
AUTHORS Kyung, K. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RP11-483A20
JOURNAL Unpublished (2001)
MEDLINE 3 (bases 1 to 175081)
PUBMED Waterston, R.H.
99063792
9847074
REFERENCE 2 (bases 1 to 175081)
AUTHORS Kyung, K. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RP11-483A20
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 175081)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 175081)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 175081)
AUTHORS Waterston, R.

TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Aug 9, 2001 this sequence version replaced gi:13877272.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@wustl.edu
 ----- Summary Statistics

 Center project name: H_NH0483A20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.regen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the right is RP11-36G9. Actual start of this clone is at base position 1 of RP11-483A20; actual end is at base position 175081 of RP11-483A20.

FEATURES

```

source
1..175081
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="4"
  /map="4"
  /clone="RP11-483A20"
  /clone_lib="RPCI-11"
repeat_region
1..3267
  /rpt_family="L1"
repeat_region
3270..3453
  /rpt_family="Alu"
repeat_region
3434..3464
  /rpt_family="(A)n"
repeat_region
3455..4161
  /rpt_family="L1"
misc_feature
7949..7956
  /note="match to EST AW006057 (NID:G59854835) wz81d04.x1"
repeat_region
10197..10489
  /rpt_family="L1"
repeat_region
10480..10501
  /rpt_family="AT-rich"
repeat_region
10495..10665
  /rpt_family="L1"
repeat_region
10731..11051
  /rpt_family="L1"
11052..11576
  /rpt_family="L1"
repeat_region
11564..11598
  /rpt_family="(A)n"
repeat_region
11642..12324
  /rpt_family="L1"
repeat_region
11769..11794
  /rpt_family="AT-rich"
repeat_region
11939..11962
  /rpt_family="AT-rich"
repeat_region
12335..12412
  /rpt_family="L1"
repeat_region
12392..12621
  /rpt_family="L1"
repeat_region
12988..13096
  /rpt_family="Alu"
repeat_region
13069..13102
  /rpt_family="(CAAA)n"
repeat_region
13325..13667
  /rpt_family="Alu"
repeat_region
13444..13484
  /rpt_family="(TAAA)n"
repeat_region
13640..13686
  /rpt_family="(GAAA)n"
repeat_region
14249..14422
  /rpt_family="MER1_type"
repeat_region
14980..15182
  /rpt_family="L1"
repeat_region
15260..15482
  /rpt_family="L1"
repeat_region
15483..15985
  /rpt_family="MER1_type"
repeat_region
15986..16105
  /rpt_family="L1"
repeat_region
16117..16251
  /rpt_family="L1"
repeat_region
16288..16356
  /rpt_family="L1"
repeat_region
16373..16545
  /rpt_family="L1"
repeat_region
16550..18773
  /rpt_family="L1"
repeat_region
18770..20218
  /rpt_family="L1"
repeat_region
19228..19346
  /rpt_family="T-rich"
repeat_region
20219..20514
  /rpt_family="Alu"
repeat_region
20515..22484
  /rpt_family="L1"
repeat_region
22584..22882
  /rpt_family="Alu"
repeat_region
22748..22768
  /rpt_family="AT-rich"
repeat_region
22883..23460
  /rpt_family="L1"
repeat_region
23461..23747
  /rpt_family="MALR"
repeat_region
23749..24514
  /rpt_family="L1"
misc_feature
24456
  /note="match to EST AW902102 (NID:G8066307)"
misc_feature
24488..24797
  /note="match to EST AW902102 (NID:G8066307)"
repeat_region
24576..24625
  /rpt_family="Mariner"
repeat_region
25484..25787
  /rpt_family="Alu"
repeat_region
25758..25787
  /rpt_family="(A)n"
repeat_region
25796..26014
  /rpt_family="MIR"
repeat_region
26657..26889

```


Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mcenhgri.nih.gov
 Akhter, N., Ayelle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 69 Row: n Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 14249445.

FEATURES

Location/Qualifiers

1..3452
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:44889 IMAGE:5574637"
 /tissue_type="Duodenum, adenocarcinoma"
 /clone_lib="NIH_MGC_88"
 /lab_host="DH105"
 /note="Vector: pCMV-SPORT6"

gene

1..3452
 /gene="FLJ14431"
 /db_xref="LocusID:84869"
 159..872

CDS

/codon_start=1
 /product="FLJ14431 protein"
 /protein_id="AAH33650.1"
 /db_xref="GI:21707067"
 /db_xref="LocusID:84869"
 /translation="MDKVCVFGSRGIGRAVLAQLMARGYRLAVLARNLEGKARAG
 DLGGDLAFSCDAKEDVDQNTPEELEKHLGRVNFVLAAGINRDLGLVRYKTEDMVS
 QLNTNLGSLTCKAARMIMIQOQSIVNGSLKNGSGSVSASRGLGVGFSR
 ALAKEVARKIRVNVVAPGVFVDMTKDLKBEHLKKNIPILGREGETIEVAHAVVFLLE
 SPYITGHVILVDGGLQLL"
 159..857

misc_feature

/note="adh short; Region: short chain dehydrogenase. This
 family contains a wide variety of dehydrogenases"
 /db_xref="CDD:pfam00106"

ORIGIN

Query Match 44.0%; Score 631; DB 9; Length 3452;
 Best Local Similarity 99.3%; Pred. No. 61e-305;
 Matches 1291; Conservative 0; Mismatches 6; Indels 3; Gaps 3;
 1 CATTAACTCTTTTATCGGGAATAGTATGATATTTCAATGTCACCTCCATTCATGTGA 2093
 2034 CATTAACTCTTTTATCGGGAATAGTATGATATTTCAATGTCACCTCCATTCATGTGA
 61 TTGGAGCTGACAGTTATTTGTGAAGCAGAGATTAAATTTATTTGAAAGTCAGTGC 120
 2094 TTGGAGCTGACAGTTATTTGTGAAGCAGAGATTAAATTTATTTGAAAGTCAGTGC 2153
 121 AAAATATTGAATAGGATATATAATAAATAACAAGTAATAACAAGTCAAGCAGTGT 180
 2154 AAAATATTGAATAGGATATATAATAAATAACAAGTAATAACAAGTCAAGCAGTGT 2213
 181 CTAATAAATAATCTGGGTTCCCTTAAATAATTTTAAATTTATC-TTGAATAGTTTTTC 239
 2214 CTAATAAATAATCTGGGTTCCCTTAAATAATTTTAAATTTATC-TTGAATAGTTTTTC 2273
 240 TTAGATTAACTCAGGATATGAAAGTCAATTAAGTGTCGTAAGTAGTATCATTTAA 299
 2274 TTAGATTAACTCAGGATATGAAAGTCAATTAAGTGTCGTAAGTAGTATCATTTAA 2333
 300 ACAAAATGTCATTAATAATGAMGA-GTGGAATATACAGAAATTTATCAGGCATTTACCAAG 358
 2334 ACAAAATGTCATTAATAATGACAGCGTGGTAATATACAGAAATTTATCAGGCATTTACCAAG 2393

QY 359 TCTAGGCACATATAGGAAATGACGACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTG 418
 Db 2394 TCTAGGCACATATAGGAAATGACGACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTG 2453
 QY 419 TAAGGTAGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGTGTSTCAATGTCT 478
 Db 2454 TAAGGTAGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGTGTSTCAATGTCT 2513
 QY 479 GGCCTTTGGCTACTCTGACTTCCSCATATGAGCAGCCCATTCAGTCTTGAAGTTTCTTCT 538
 Db 2514 GGCCTTTGGCTACTCTGACTTCCSCATATGAGCAGCCCATTCAGTCTTGAAGTTTCTTCT 2573
 QY 539 CTGACACACCTTATGCTCTGAATCATGAGCAGGCTGATTCAATTTGGTGAATTTGGGTAGA 598
 Db 2574 CTGACACACCTTATGCTCTGAATCATGAGCAGGCTGATTCAATTTGGTGAATTTGGGTAGA 2633
 QY 599 AAGCAGTATGTTTGTCTGACATTAAGATGTAGTGTATAGATAGGTTTACGCTTTAAGTGT 658
 Db 2634 AAGCAGTATGTTTGTCTGACATTAAGATGTAGTGTATAGATAGGTTTACGCTTTAAGTGT 2693
 QY 659 ATGTTTTTATACCTTAAATAAAGAAATATAACCTTTTAAAGCTATTCACCTCTCCCCCA 718
 Db 2694 ATGTTTTTATACCTTAAATAAAGAAATATAACCTTTTAAAGCTATTCACCTCTCCCCCA 2753
 QY 719 GCCTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCA 778
 Db 2754 GCCTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCA 2813
 QY 779 CTGCTCCACTCCAGGTGAATCCGCCCACTCCCACTGACCTAGTAGAGTAATTTGTAATTAAT 838
 Db 2814 CTGCTCCACTCCAGGTGAATCCGCCCACTCCCACTGACCTAGTAGAGTAATTTGTAATTAAT 2873
 QY 839 ACTTACTCTTATTTCTGAAATCAGTTGTGAACCTGTTCCTTATGTTTCAAGTGTAAAGA 898
 Db 2874 ACTTACTCTTATTTCTGAAATCAGTTGTGAACCTGTTCCTTATGTTTCAAGTGTAAAGA 2933
 QY 899 ACCTCMGTGAATTCATTTTTTAAATCTGCTATTTCTGAGAAGCATTTGAATGATCTTAA 958
 Db 2934 ACCTCAGTGAATTCATTTTTTAAATCTGCTATTTCTGAGAAGCATTTGAATGATCTTAA 2993
 QY 959 CAAGAAGACTCATCTGCTAGTGTTCCTGACTCTCTATGAGCCCAATAAGGTTCTGTGCT 1018
 Db 2994 CAAGAAGACTCATCTGCTAGTGTTCCTGACTCTCTATGAG-CCCAATAAGGTTCTGTGCT 3052
 QY 1019 TAGCATTAACAAATAAAGTTTATAGGTAAGCAATGTATTAAATTTTTTTTGCATGGA 1078
 Db 3053 TAGCATTAACAAATAAAGTTTATAGGTAAGCAATGTATTAAATTTTTTTTGCATGGA 3112
 QY 1079 GGGCTTTAAATTTGTGCTCTTTTCAATTTTATTCATATTCAATTTATGTTTGTGTAAC 1138
 Db 3113 GGGCTTTAAATTTGTGCTCTTTTCAATTTTATTCATATTCAATTTATGTTTGTGTAAC 3172
 QY 1139 TGCCTTTTAGGAGATTAATATATGTTATAAATAGTTTGGGGGAAATAATGTCAAA 1198
 Db 3173 TGCCTTTTAGGAGATTAATATATGTTATAAATAGTTTGGGGGAAATAATGTCAAA 3232
 QY 1199 GAGGATAATTTAAATTTACGTGCTTCTGTTATTTCAGAAATTAAGAGAGACGCTGCA 1258
 Db 3233 GAGGATAATTTAAATTTACGTGCTTCTGTTATTTCAGAAATTAAGAGAGACGCTGCA 3292
 QY 1259 TATTCAGAGTTGTACCTTAAACATTTGGTGAACATTTTTT 1298
 Db 3293 TATTCAGAGTTGTACCTTAAACATTTGGTGAACATTTTTT 3332
 RESULT 6
 AC117374 281662 bp DNA linear HTG 02-MAY-2002
 LOCUS
 DEFINITION Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
 PROGRESS ***, 48 unordered pieces.
 AC117374 AC011694
 AC117374.1 GI:20127352

HTG: HTGS PHASE1.
 Homo sapiens (human)
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 1 (bases 1 to 281662)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbiana,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulseg,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
 Ogayune,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
 Sodergren,E., Sonaiker,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Teifrod,B., Thomas,N., Thomas,S.,
 Umani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 281662)
 Worley,K.C.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (10-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 281662)
 Worley,K.C.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 28, 2002 this sequence version replaced gi:10045385.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HCNB
 Center clone name: RP11-19D19
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990129

Consensus quality: 273063 bases at least Q40
 Consensus quality: 294336 bases at least Q30
 Consensus quality: 306942 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 48 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 2464: contig of 2464 bp in length
 * 2465: gap of unknown length
 * 2564: contig of 2039 bp in length
 * 2565: contig of 2039 bp in length
 * 4603: gap of unknown length
 * 4604: contig of 2410 bp in length
 * 4703: contig of 2410 bp in length
 * 7113: gap of unknown length
 * 7213: contig of 2304 bp in length
 * 9517: gap of unknown length
 * 9617: gap of unknown length
 * 9618: contig of 2490 bp in length
 * 12107: contig of 2490 bp in length
 * 12207: gap of unknown length
 * 15014: contig of 2807 bp in length
 * 15015: gap of unknown length
 * 15115: contig of 2717 bp in length
 * 17831: gap of unknown length
 * 17931: contig of 2549 bp in length
 * 20480: gap of unknown length
 * 20580: gap of unknown length
 * 20481: contig of 2120 bp in length
 * 22700: gap of unknown length
 * 22800: contig of 2109 bp in length
 * 24903: gap of unknown length
 * 25003: gap of unknown length
 * 27313: contig of 2304 bp in length
 * 27413: gap of unknown length
 * 29798: contig of 2385 bp in length
 * 29898: gap of unknown length
 * 33466: contig of 3568 bp in length
 * 33566: gap of unknown length
 * 36679: contig of 3113 bp in length
 * 36779: gap of unknown length
 * 36780: contig of 3693 bp in length
 * 40472: gap of unknown length
 * 40572: gap of unknown length
 * 42683: contig of 2111 bp in length
 * 42684: gap of unknown length
 * 46319: contig of 3536 bp in length
 * 46320: gap of unknown length
 * 46419: gap of unknown length
 * 49507: contig of 3088 bp in length
 * 49607: gap of unknown length
 * 53212: contig of 3605 bp in length
 * 53312: gap of unknown length
 * 57267: contig of 3955 bp in length
 * 57367: gap of unknown length
 * 60444: contig of 3077 bp in length
 * 60544: gap of unknown length
 * 65853: contig of 5315 bp in length
 * 65953: gap of unknown length
 * 69038: contig of 3079 bp in length
 * 69138: gap of unknown length
 * 73445: contig of 4307 bp in length
 * 73446: gap of unknown length
 * 73545: contig of 4562 bp in length
 * 78107: gap of unknown length
 * 78207: gap of unknown length
 * 82504: contig of 4297 bp in length
 * 82604: gap of unknown length
 * 86976: contig of 4372 bp in length
 * 87076: gap of unknown length
 * 91315: contig of 4239 bp in length
 * 91415: gap of unknown length
 * 95084: contig of 3669 bp in length
 * 95184: gap of unknown length
 * 102348: contig of 7164 bp in length
 * 95185

```

* 102349 102448: gap of unknown length
* 102449 105973: contig of 3525 bp in length
* 105974 106073: gap of unknown length
* 106074 112891: contig of 6818 bp in length
* 112892 112991: gap of unknown length
* 112992 116221: contig of 3330 bp in length
* 116222 116321: gap of unknown length
* 116322 117998: contig of 3477 bp in length
* 117999 119899: gap of unknown length
* 119900 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140737: gap of unknown length
* 140738 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155057: contig of 9416 bp in length
* 155058 155157: gap of unknown length
* 155158 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174892: contig of 10435 bp in length
* 174893 174992: gap of unknown length
* 174993 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 216536: gap of unknown length
* 216537 232916: contig of 16380 bp in length
* 232917 233016: gap of unknown length
* 233017 255437: contig of 22421 bp in length
* 255438 255537: gap of unknown length
* 255538 281662: contig of 26125 bp in length.
FEATURES
    Location/Qualifiers
        source
            1..281662
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="12"
            /clone="RP11-19D19"
ORIGIN
    Query Match      27.9%; Score 400; DB 2; Length 281662;
    Best Local Similarity 100.0%; Pred. No. 1.9e-189;
    Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1000 CCCATAAGGGTTCGTCTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAT 1059
DB 34088 CCCATAAGGGTTCGTCTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAT 34147

QY 1060 TAAATTTTTTTTGCATGGAGGGCTTTAAATTTCTGCTCTTTTTCATATTTTATTCATAT 1119
DB 34148 TAAATTTTTTTTGCATGGAGGGCTTTAAATTTCTGCTCTTTTTCATATTTTATTCATAT 34207

QY 1120 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATGTTTATAAATTAGTTTGT 1179
DB 34208 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATGTTTATAAATTAGTTTGT 34267

QY 1180 GGGGGAATAATTGCAAGAGGATAATTAAATTTAGTGTCTCTGTTTATTCAGAAATAAA 1239
DB 34268 GGGGGAATAATTGCAAGAGGATAATTAAATTTAGTGTCTCTGTTTATTCAGAAATAAA 34327

QY 1240 GAGAGAAGACTACGCTGCATATTCAAGAGTTGTACCTTTAACTTTGAGAAACATTTTTC 1299
DB 34328 GAGAGAAGACTACGCTGCATATTCAAGAGTTGTACCTTTAACTTTGAGAAACATTTTTC 34387

QY 1300 TAAGATTTTCAAAGGAATATGTTAAATTTAGAAATTCATAACCACTGTCTCTAACTTGGT 1359
DB 34388 TAAGATTTTCAAAGGAATATGTTAAATTTAGAAATTCATAACCACTGTCTCTAACTTGGT 34447

```

```

QY 1360 AACAAACTGTTCTTAATAAAGTATTTAATGATTTTAA 1399
DB 34448 AACAAACTGTTCTTAATAAAGTATTTAATGATTTTAA 34487

RESULT 7
AX408738/c
LOCUS AX408738 448 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1385 from Patent WO0229103.
ACCESSION AX408738
VERSION AX408738.1 GI:21441443
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1. Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
    Gene expression profiles in liver cancer
    Patent: WO 0229103-A 1385 11-APR-2002;
    GENE LOGIC INC (US)
FEATURES
    Location/Qualifiers
        source
            1..448
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="EMBL/GenBank Accession No. AA491000"
ORIGIN
    Query Match      24.1%; Score 345; DB 6; Length 448;
    Best Local Similarity 100.0%; Pred. No. 1.3e-161;
    Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1000 CCCATAAGGGTTCGTCTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAT 1059
DB 414 CCCATAAGGGTTCGTCTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAT 355

QY 1060 TAAATTTTTTTTGCATGGAGGGCTTTAAATTTCTGCTCTTTTTCATATTTTATTCATAT 1119
DB 354 TAAATTTTTTTTGCATGGAGGGCTTTAAATTTCTGCTCTTTTTCATATTTTATTCATAT 295

QY 1120 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATGTTTATAAATTAGTTTGT 1179
DB 294 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATGTTTATAAATTAGTTTGT 235

QY 1180 GGGGGAATAATTGCAAGAGGATAATTAAATTTAGTGTCTCTGTTTATTCAGAAATAAA 1239
DB 234 GGGGGAATAATTGCAAGAGGATAATTAAATTTAGTGTCTCTGTTTATTCAGAAATAAA 175

QY 1240 GAGAGAAGACTACGCTGCATATTCAAGAGTTGTACCTTTAACTTTGAGAAACATTTTTC 1299
DB 174 GAGAGAAGACTACGCTGCATATTCAAGAGTTGTACCTTTAACTTTGAGAAACATTTTTC 115

QY 1300 TAAGATTTTCAAAGGAATATGTTAAATTTAGAAATTCATAACCA 1344
DB 114 TAAGATTTTCAAAGGAATATGTTAAATTTAGAAATTCATAACCA 70

```

```

RESULT 8
AC117374/c
LOCUS AC117374 281662 bp DNA linear HTG 02-MAY-2002
DEFINITION Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
            PROGRESS ***, 48 unordered pieces.
ACCESSION AC117374 AC011694
VERSION AC117374.1 GI:20127352
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1. (bases 1 to 281662)
    Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
    Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,

```

Barbaria, J., Benton, J., Bimage, K., Blaukenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H.,
 Lorzado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mashiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oreguena, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 281662)
 Worley, K.C.

Submitted (10-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 281662)
 Worley, K.C.

Direct Submission
 Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Apr 28, 2002 this sequence version replaced gi:10045385.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: HCNB
 Center clone name: RP11-19D19
 Summary Statistics

----- Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 273063 bases at least Q40
 Consensus quality: 294336 bases at least Q30
 Consensus quality: 306942 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 48 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 2464: contig of 2464 bp in length
 2465: gap of unknown length
 2564: contig of 2039 bp in length
 2565: gap of unknown length
 4603: contig of 2410 bp in length
 4703: gap of unknown length
 7113: contig of 2304 bp in length
 7214: gap of unknown length
 9518: contig of 2490 bp in length
 9618: gap of unknown length
 12107: contig of 2807 bp in length
 12208: gap of unknown length
 15015: contig of 2717 bp in length
 15114: gap of unknown length
 17831: contig of 2549 bp in length
 17932: gap of unknown length
 20480: contig of 2120 bp in length
 20581: gap of unknown length
 22701: contig of 2109 bp in length
 22801: gap of unknown length
 24910: contig of 2304 bp in length
 25010: gap of unknown length
 27314: contig of 2385 bp in length
 27414: gap of unknown length
 29799: contig of 3568 bp in length
 29899: gap of unknown length
 33567: contig of 3113 bp in length
 33679: gap of unknown length
 36780: contig of 3693 bp in length
 40473: gap of unknown length
 40573: contig of 2111 bp in length
 42684: gap of unknown length
 42784: contig of 3536 bp in length
 46320: gap of unknown length
 46420: contig of 3088 bp in length
 49508: gap of unknown length
 49608: contig of 3605 bp in length
 53213: gap of unknown length
 53313: contig of 3955 bp in length
 57268: gap of unknown length
 57368: contig of 3077 bp in length
 60445: gap of unknown length
 60545: contig of 5315 bp in length
 65860: gap of unknown length
 65960: contig of 3079 bp in length
 69039: gap of unknown length
 69139: contig of 4307 bp in length
 73446: gap of unknown length
 73546: contig of 4562 bp in length
 78108: gap of unknown length
 78208: contig of 4297 bp in length
 82505: gap of unknown length
 82605: contig of 4372 bp in length
 86977: gap of unknown length
 91316: contig of 4239 bp in length
 91416: gap of unknown length
 95085: contig of 3669 bp in length
 95185: gap of unknown length
 102349: contig of 7164 bp in length
 102449: gap of unknown length
 105973: contig of 3525 bp in length
 105974: gap of unknown length
 112892: contig of 6818 bp in length
 112991: gap of unknown length
 116221: contig of 3230 bp in length
 116222: gap of unknown length
 119798: contig of 3477 bp in length

```

* 119799 119898: gap of unknown length
* 119899 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140737: gap of unknown length
* 140738 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155057: contig of 9416 bp in length
* 155058 155157: gap of unknown length
* 155158 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174892: contig of 10435 bp in length
* 174893 174992: gap of unknown length
* 174993 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 216536: gap of unknown length
* 216537 232915: contig of 16380 bp in length
* 232916 233015: gap of unknown length
* 233017 255438: contig of 22421 bp in length
* 255438 255537: gap of unknown length
* 255537 281662: contig of 26125 bp in length.
FEATURES
source
    1. .281662
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="12"
    /clone="RP11-19D19"
ORIGIN
    Query Match      13.08; Score 187; DB 2; Length 281662;
    Best Local Similarity 100.0%; Pred. No. 9.2e-83;
    Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41  TGTCACTCCATTGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTAAAT 100
Db 146186 TGTCACTCCATTGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTAAAT 146127

QY 101 TTTATATTGAAGTCAGTCAAAATTTATGAATAGGATATATACTATAATAACAAAGTAATA 160
Db 146126 TTTATATTGAAGTCAGTCAAAATTTATGAATAGGATATATACTATAATAACAAAGTAATA 146067

QY 161 ACAAAAGTCAAAGCAGTCTTAATAATAAAATTCGGTTCCTTAAAAATTTTAAAT 220
Db 146066 ACAAAAGTCAAAGCAGTCTTAATAATAAAATTCGGTTCCTTAAAAATTTTAAAT 146007

QY 221 TTATCTT 227
Db 146006 TTATCTT 146000

RESULT 9
AX898296
LOCUS
    Sequence 14159 from Patent EP1033401.
DEFINITION
AX898296
ACCESSION
VERSION
AX898296.1 GI:40053209
KEYWORDS
SOURCE
    Homo sapiens (human)
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
    Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
    Expressed sequence tags and encoded human proteins
    Patent: EP 1033401-A 14159 06-SEP-2000;

```

```

Genset (FR)
FEATURES
source
    1. .395
    /organism="Homo sapiens"
    /mol_type="unassigned DNA"
    /db_xref="taxon:9606"
ORIGIN
    Query Match      9.9%; Score 142; DB 6; Length 395;
    Best Local Similarity 100.0%; Pred. No. 5.9e-60;
    Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58  TGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTAAATTTATTTGAAAGTCAG 117
Db 1   TGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTAAATTTATTTGAAAGTCAG 60

QY 118 TGC AAAATTTATGATAGGATATATACTATAATAACAAGTAATAACAAAGTCAAGCAGT 177
Db 61  TGC AAAATTTATGATAGGATATATACTATAATAACAAGTAATAACAAAGTCAAGCAGT 120

QY 178 GTTCTAAATAAAAATTCGGGT 199
Db 121 GTTCTAAATAAAAATTCGGGT 142

RESULT 10
BD033829
LOCUS
    Sequence tag and encoded human protein.
DEFINITION
BD033829
ACCESSION
BD033829.1 GI:22575571
KEYWORDS
    JP 2001269182-A/10075.
SOURCE
    Homo sapiens (human)
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
    (bases 1 to 395)
    Edwards,J.B.D.M., Duclair,E. and Jordan,J.Y.
    Sequence tag and encoded human protein
    Patent: JP 2001269182-A 10075 02-OCT-2001;
    GENSSET
COMMENT
    OS Homo sapiens (human)
    PN JP 2001269182-A/10075
    PD 02-OCT-2001
    PF 24-FEB-2000 JP 2000118773
    PR 26-FEB-1999 US 60/122487
    PI JEAN RAPUIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
    PI JORDAN
    PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21,PC
    C12N5/10,
    PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00,PC
    G06F15/40
CC
FH Key Location/Qualifiers.
FEATURES
source
    1. .395
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
ORIGIN
    Query Match      9.9%; Score 142; DB 6; Length 395;
    Best Local Similarity 100.0%; Pred. No. 5.9e-60;
    Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58  TGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTAAATTTATTTGAAAGTCAG 117
Db 1   TGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTAAATTTATTTGAAAGTCAG 60

QY 118 TGC AAAATTTATGATAGGATATATACTATAATAACAAGTAATAACAAAGTCAAGCAGT 177
Db 61  TGC AAAATTTATGATAGGATATATACTATAATAACAAGTAATAACAAAGTCAAGCAGT 120

```

```

Qy 178 GTTCTAATAAATAATTCGGGT 199
Db 121 GTTCTAATAAATAATTCGGGT 142

RESULT 11
BC063167 1648 bp mRNA linear ROD 03-DEC-2003
LOCUS Rattus norvegicus ADP-ribosylation factor 4, mRNA (cDNA clone
DEFINITION MGC:72808 IMAGE:6921397), complete cds.
ACCESSION BC063167
VERSION BC063167.1 GI:38648903
KEYWORDS MGC.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1648)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buéto, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Igoe, L., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1648)
Straussberg, R.
Direct Submission
Submitted (02-DEC-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: John C. Marshall, M.D., Ph.D
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAL Plate: 52 Row: 0 Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 13162342.
Location/Qualifiers
1. .1648

FEATURES
source
Location/Qualifiers
1. .8281
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="3"
/clone="DKFZp434P0721"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"

/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="MGC:72808 IMAGE:6921397"
/tissue_type="Pituitary gland, anterior, rat"
/clone_lib="NICHDR_Pit1"
/lab_host="DH10B"
/note="vector: pDNR-LIB"
1. .1648
/gene="Arf4"
/db_xref="LocusID:79120"
/db_xref="RGD:621275"
178. .720
/codon_start=1
/product="ADP-ribosylation factor 4"
/protein_id="AAH63167.1"
/db_xref="GI:38648904"
/db_xref="LocusID:79120"
/translation="MGLTISLSFRLPKQKQIRLMLVGLDAAGKTTLYKLGLGEIVT
TPTIGNVETVEYKNICFTVWDVGGQDKIRPLWRHYFQNTQGLIFVVDNSNDRERIOE
GAAVLQWLLDELDQAVLLLEFANKQDLPNMAISEMTDKLGSLNRTWYVQATCA
TQGTGLYEGLDWLSNELSKR"
178. .717
/note="ARF; Region: ARF-like small GTPases"
/db_xref="CDD:smart00177"

Query Match 3.3%; Score 47; DB 10; Length 1648;
Best Local Similarity 100.0%; Pred. No. 2e-12;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1388 AATGATTTTAAAAA 1434
Db 1556 AATGATTTTAAAAA 1602

RESULT 12
HSM801490 8281 bp mRNA linear PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFZp434P0721 (from clone DKFZp434P0721);
DEFINITION partial cds.
ACCESSION AL133623
VERSION AL133623.1 GI:6599261
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 8281)
Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
Direct Submission
Submitted (15-DEC-1999) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp434P0721) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES
source
Location/Qualifiers
1. .8281
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="3"
/clone="DKFZp434P0721"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"

```

```

gene
  /dev_stage="adult"
  1. .8281
  /gene="DKEZp434P0721"
CDS
  <1. .3481
  /gene="DKEZp434P0721"
  /note="strong similarity to mouse 5'-3' exonuclease"
  /codon_start=2
  /product="hypothetical protein"
  /protein_id="CAB63749.1"
  /db_xref="GI:6599262"
  /db_xref="GOA:Q9UF39"
  /db_xref="SPTRMBL:O9UF39"
  /translation="TNEDSPLIIEYDPDKTDLNGKQWEAVVLPIFIDERLLEAME
TCNHSKKERKRNQSECLWCYDTEFIYSPMPKEKFAIERCCTRYKIIISLDAW
RVINDKNKTRIDKALYFCGPTFLKIRHKFKKSGVQVFOQSGRGENMELIIVD
AESDELTENVASVGLKSVFVNMPLHEEARVAVSDGETKFLEPPGQKLYSGRT
APPSKVHLGDQKSNWAKEVOGISEHYLRKGIINETSAVVYAOALLTKRQIINON
GEVLRKQWQKQVFPVYQTIIVKDIRAFDSRFENIKTLDDLFLPLRSVMFLSTPYGC
TGEVQSGDVTIGRIRVIFSIPECPNLDALIQNHKYSIKYNPGLVIALRGLVSGYL
VSFTGSIFIGRGRNRNPHGDKANVLNFKNKEEVEPGYTKKGVSEWMSAAEQ
LLAAYLERAPELFSYIAKNSQEDVYEDDIWPGENENGAEKVQEIITWLKHFPVSTLS
RSSCDLIILDAALIVEKEIEVEKQKRNKKVTVTKPHLLYRPLEQHGVIPOKDA
EFCFLFORVNVNRENFSPVGLRTIIGIKANREADVLPVLFDPEEPGLTIRCSPG
RGVRLPTSLAVNLSHGSRSTGNKLTALVQPAPVHOHSSSSSVSGHLGALNHSPQ
SLVFPVQVPTKDDFCFNWQSLQSGSKMYFQPTIOEKGAVLPOEISQVQNHKSGF
NDSVKIQQRKHPHRRKBEKSPKAEQSKWQSNQKPNQSGNFIENFLASLNISKENEV
QSHGHEPPSEELHSPQSFAMKGTMLKEILKIDGNTVDHKNKIKQIANEIVPSNLR
RDYGLFSPQKQKLLASYNKPHSANEYHNQSMNMCWPAPSQIPVPVPTVELSR
ICSLGMPQDFSLRMPQMTVCQVKSNGLLVHGPOCHSENAKEKAALFALQQLG
SLGMNTPLRSQVAPVSPVPGTIPPAFPPTANIMPSSHLFGSPWPCSPVPVPGK
PEHTILYSGTMWAGGIPGVHNOQFIPLOVTKRVANKNFENKEAQSSATPVQISQ
PDSNVLVFPRESSASLSSPLAQPASSQFVETASQGHSHHKSTPISSRRKSR
KLAVNFGVSKPSE"
  8205. .8210
  /gene="DKEZp434P0721"
  8227
  /gene="DKEZp434P0721"
ORIGIN
  polyA_signal
  polyA_site

```

```

Query Match      3.3%; Score 47; DB 9; Length 8281;
Best Local Similarity 100.0%; Pred. No. 1.7e-12;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1388 AATGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
      |||||||
Db 8219 AATGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 8265

```

RESULT 13

```

AK113242
LOCUS      AK113242      1378 bp      mRNA      linear      INV 30-NOV-2002
DEFINITION Ciona intestinalis cDNA, clone:ciad019c13, full insert sequence.
ACCESSION  AK113242
VERSION     AK113242.1  GI:23577212
KEYWORDS   F11 CDNA.
SOURCE      Ciona intestinalis
ORGANISM    Ciona intestinalis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.

```

```

REFERENCE   1
AUTHORS     Satou, Y., Yamada, L., Mochizuki, Y., Takatori, N., Kawashima, T.,
            Sasaki, A., Hamaguchi, M., Awazu, S., Yagi, K., Sasakura, Y.,
            Nakayama, A., Ishikawa, H., Inaba, K. and Satoh, N.
            A cDNA resource from the basal chordate Ciona intestinalis
            Genesis 33 (4), 153-154 (2002)
MEDLINE     22191024
PUBMED      12203911

```

```

REFERENCE   2 (bases 1 to 1378)
AUTHORS     Satou, Y. and Satoh, N.
TITLE       Direct Submission
JOURNAL     Submitted (04-OCT-2002) Nori Satoh, Kyoto University, Department of
            Zoology; Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            (E-mail:satoh@ascidian.zool.kyoto-u.ac.jp, Tel:81-75-753-4095,

```

```

COMMENT     Fax:81-75-705-1113)
            Ciona intestinalis cDNA Project (URL:
            http://ghost.zool.kyoto-u.ac.jp/indexr1.html).
FEATURES
  source
    1. .1378
    /organism="Ciona intestinalis"
    /mol_type="mRNA"
    /db_xref="taxon:7719"
    /clone="ciad019c13"
ORIGIN

```

```

Query Match      3.2%; Score 46; DB 3; Length 1378;
Best Local Similarity 100.0%; Pred. No. 6.3e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1389 ATGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
      |||||||
Db 1328 ATGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1373

```

RESULT 14

```

PFMAL13_16
WPCOMMENT
  Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509

```

Fragment Name	Begin	End
PFMAL13_00	1	110000
PFMAL13_01	100001	210000
PFMAL13_02	200001	310000
PFMAL13_03	300001	410000
PFMAL13_04	400001	510000
PFMAL13_05	500001	610000
PFMAL13_06	600001	710000
PFMAL13_07	700001	810000
PFMAL13_08	800001	910000
PFMAL13_09	900001	1010000
PFMAL13_10	1000001	1110000
PFMAL13_11	1100001	1210000
PFMAL13_12	1200001	1310000
PFMAL13_13	1300001	1410000
PFMAL13_14	1400001	1510000
PFMAL13_15	1500001	1610000
PFMAL13_16	1600001	1710000
PFMAL13_17	1700001	1810000
PFMAL13_18	1800001	1910000
PFMAL13_19	1900001	2010000
PFMAL13_20	2000001	2110000
PFMAL13_21	2100001	2210000
PFMAL13_22	2200001	2310000
PFMAL13_23	2300001	2410000
PFMAL13_24	2400001	2510000
PFMAL13_25	2500001	2610000
PFMAL13_26	2600001	2710000
PFMAL13_27	2700001	2732359

Continuation (17 of 28) of PFMAL13 from base 1600001 (AL844509 Plasmodium falciparum 3D)

```

Query Match      3.2%; Score 46; DB 2; Length 110000;
Best Local Similarity 100.0%; Pred. No. 4.1e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1389 ATGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
      |||||||
Db 20114 ATGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 20159

```

RESULT 15

```

BC009267
LOCUS      BC009267      1648 bp      mRNA      linear      PRI 16-SEP-2003
DEFINITION Homo sapiens lamin B2, mRNA (cDNA clone IMAGE:4053047), partial
            cds.
ACCESSION  BC009267
VERSION     BC009267.2  GI:33872541
KEYWORDS
SOURCE      Homo sapiens (human)

```

Search completed: July 20, 2004, 23:04:25
Job time : 5505 secs

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 1648)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heien,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Sapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,D.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalios,D.E.,
Schnerch,A., Schein,J.B., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
MEDLINE
12477932
PUBMED
2 (bases 1 to 1648)
Strausberg,R.
Direct Submission
REFERENCE
TITLE
JOURNAL
Submitted (08-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT
NIH-MGC Project URL: http://mgc.nci.nih.gov
On Aug 19, 2003 this sequence version replaced gi:14349378.
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadanesystemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 25 Row: 0 Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14249359.
FEATURES
source
location/Qualifiers
1..1648
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4033047"
/tissue_type="uterus, leiomyosarcoma"
/clone_lib="NIH MGC 46"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
ORIGIN
Query Match 3.1%; Score 45; DB 9; Length 1648;
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1390 TGATTTTAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
|||||
DB 1604 TGATTTTAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1648

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 18:53:19 ; Search time 583 Seconds
(without alignments)
10449.256 Million cell updates/sec

Title: US-10-799-747-12

Perfect score: 1434

Sequence: 1 cattaactctttttatcg.....aaaaaaaaaaaaaaaaaa 1434

Scoring table:

OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq_29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002s.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1434	100.0	1434	7 ADA39771	Human sec
2	1434	100.0	1434	7 ACC50424	Human sec
3	1383	96.4	1434	2 AAX37452	Human sec
C 4	345	24.1	448	6 ABN94887	Gene #138
5	145	10.1	415	3 AAH30357	Human col
6	142	9.9	395	3 AAC10084	Human sec
7	127	8.9	336	2 AAT25136	Human gen
C 8	45	3.1	493	9 ADE81840	Human gen
9	45	3.1	1231	4 AAH25190	Arabidops
10	45	3.1	1348	2 AAZ24411	Nucleotid
11	45	3.1	2552	6 ABL89915	Human bla
12	45	3.1	3470	5 ABV25066	Human pol
13	44	3.1	1730	3 AAA87997	Human pro
14	43	3.0	222	2 AAQ46071	Torenia h
15	43	3.0	222	4 AAC86777	Sequence
C 16	43	3.0	271	4 AAL16362	Downstrea
17	43	3.0	306	6 ABZ08725	Human bre
C 18	43	3.0	337	5 ABV5706	Human leu
19	43	3.0	337	6 ABQ86019	Human pro
C 20	43	3.0	339	4 AAI86180	Arabidops
C 21	43	3.0	380	5 ABAL3116	Human pol
22	43	3.0	383	6 ABL01579	Human ner
C 23	43	3.0	384	4 AAH70111	Human sec
					Human cer

24	43	3.0	399	2 AAT13044	Cotton fi
25	43	3.0	399	2 AAT30261	Cotton fi
C 26	43	3.0	399	2 AAT62620	Cotton fi
27	43	3.0	399	2 AAT70051	Cotton fi
C 28	43	3.0	399	3 AAZ35555	CDNA sequ
29	43	3.0	401	4 AAI88370	Human pol
30	43	3.0	402	4 AAI85111	Human pol
31	43	3.0	402	8 ACH18461	Human adu
32	43	3.0	418	4 AAI88540	Human pol
33	43	3.0	429	5 ABV49095	Human pro
C 34	43	3.0	480	6 ABZ08201	Human leu
C 35	43	3.0	528	6 ABZ08195	Human leu
C 36	43	3.0	556	5 ABV40163	Human pro
C 37	43	3.0	556	5 ABV40063	Human pro
C 38	43	3.0	556	5 ABV42105	Human pro
C 39	43	3.0	556	5 ABV43601	Human pro
C 40	43	3.0	638	6 ABZ78098	Human bre
41	43	3.0	681	5 ABV19321	Human pro
42	43	3.0	785	3 AAC77913	Human can
43	43	3.0	916	2 AAT65651	CDNA enco
44	43	3.0	916	2 AAV62000	R. prolif
45	43	3.0	944	4 AAH32549	Human sec

ALIGNMENTS

RESULT 1

ADA39771

ID ADA39771 standard; cDNA; 1434 BP.

XX AC ADA39771;

XX DT 20-NOV-2003 (first entry)

DE Human secreted protein encoding cDNA.

XX KW Human; secreted protein; cancer; hyperproliferative disorder;

KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;

KW anaemia; allergic reaction; asthma; cardiovascular disorder;

KW wound healing; cytostatic; immunosuppressive; neutropenic; neuroprotective;

KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;

XX KW vulnery; cardiant; gene therapy; ss.

XX OS Homo sapiens.

XX XX WO2002102993-A2.

XX PD 27-DEC-2002.

XX 19-MAR-2002; 2002WO-US008123.

XX 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-175238/17.

XX New human secreted proteins and nucleic acid molecules, useful for

PT preparing a diagnostic or pharmaceutical composition for diagnosing,

PT preventing or treating cancer or other hyperproliferative disorder,

XX asthma, allergies or AIDS.

PS Claim 9; SEQ ID NO 153; 3205pp; English.

XX The invention relates to novel genes ADA39629-ADA40565 and proteins

CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,

CC treating or ameliorating medical conditions e.g. by protein or gene

CC therapy. The polypeptides, nucleic acid molecules, antibodies or their

CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Seq Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Query Match 100.0%; Score 1434; DB 7; Length 1434;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAAGTGCACCTCCATTGATGA 60
Db |||||||
Qy 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAAGTGCACCTCCATTGATGA 60
Db |||||||
Qy 61 TTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAAATTTTATATTGAAAGTCAGTGC 120
Db |||||||
Qy 61 TTTGGAGCTGACAGTTATTTGTGTAAGCAGAGATTTAAATTTTATATTGAAAGTCAGTGC 120
Db |||||||
Qy 121 AAAATTTAATGATAGATATATAATAATAACAAGTAAATACAAAAGTCAAGCAGTGT 180
Db |||||||
Qy 121 AAAATTTAATGATAGATATATAATAATAACAAGTAAATACAAAAGTCAAGCAGTGT 180
Db |||||||
Qy 181 CTAATAAATAATCTGGGTTCTTAAATAATTTTAAATTTATCTGAAATAGTTTCT 240
Db |||||||
Qy 181 CTAATAAATAATCTGGGTTCTTAAATAATTTTAAATTTATCTGAAATAGTTTCT 240
Db |||||||
Qy 241 TAGATTAATCTCAGGATATGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTAAA 300
Db |||||||
Qy 241 TAGATTAATCTCAGGATATGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTAAA 300
Db |||||||
Qy 301 CAAATTTGCTATTAAATCGACAGTGGTAAATATACAGAAATTTATCAGGCATTACCAAGTC 360
Db |||||||
Qy 301 CAAATTTGCTATTAAATCGACAGTGGTAAATATACAGAAATTTATCAGGCATTACCAAGTC 360
Db |||||||
Qy 361 TAGGCACATATAGGAATGACAGCACTCAGATGGTTTCAATGTAGTAGTTGATGCTTGTA 420
Db |||||||
Qy 361 TAGGCACATATAGGAATGACAGCACTCAGATGGTTTCAATGTAGTAGTTGATGCTTGTA 420
Db |||||||
Qy 421 AGGTAGGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGCTGTSTCAATGTCTGG 480
Db |||||||
Qy 421 AGGTAGGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGCTGTSTCAATGTCTGG 480
Db |||||||
Qy 481 CCTTTGGCTACTGTACTTCCSCATTATGGCAGCCATTGAGTCTTGTGTTTCTTCTCT 540
Db |||||||
Qy 481 CCTTTGGCTACTGTACTTCCSCATTATGGCAGCCATTGAGTCTTGTGTTTCTTCTCT 540
Db |||||||
Qy 541 GGACACCTTATGCTCTGAAATCATGAGGAGGCTGATTCAATTTGGTGAATTTGGGTAGAAA 600
Db |||||||
Qy 541 GGACACCTTATGCTCTGAAATCATGAGGAGGCTGATTCAATTTGGTGAATTTGGGTAGAAA 600
Db |||||||
Qy 601 GCAGTATGTTTGTCTGACATTAAGATGTAGTTATAGTAGTTTAGCTTTAGTGTAT 660
Db |||||||

Db 601 GCAGTATGTTTGTCTGACATTAAGATCTAGGTATATAGATAGTGTATAGCTTTAAGTGAT 660
Qy 661 GTTTTATATCTTTAAATAAAGAAATATAACCTTTTAAAGCTATTCCACTCTCTCCCCGAGC 720
Db |||||||
Qy 661 GTTTTATATCTTTAAATAAAGAAATATAACCTTTTAAAGCTATTCCACTCTCTCCCCGAGC 720
Db |||||||
Qy 721 CTATCTCAAACTGCTGGAATATATGGAGAGATCTTGAAGAAGTAAATAAATTAACCTTCACT 780
Db |||||||
Qy 721 CTATCTCAAACTGCTGGAATATATGGAGAGATCTTGAAGAAGTAAATAAATTAACCTTCACT 780
Db |||||||
Qy 781 GCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTAATAC 840
Db |||||||
Qy 781 GCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTAATAC 840
Db |||||||
Qy 841 TTACCTTCTATTTCTGAATTCAGTTGTGAACCTGTGCTTATGTTCAGARGTTTAAGAAC 900
Db |||||||
Qy 841 TTACCTTCTATTTCTGAATTCAGTTGTGAACCTGTGCTTATGTTCAGARGTTTAAGAAC 900
Db |||||||
Qy 901 CTCMGTCGAATTCATTTTAAATCTGCTATTCTGAGAAGCATTAATGTAATCTTAACA 960
Db |||||||
Qy 901 CTCMGTCGAATTCATTTTAAATCTGCTATTCTGAGAAGCATTAATGTAATCTTAACA 960
Db |||||||
Qy 961 AGAAGACTCATCTGCTAGTCTGTTGCTGACCTCTATGAGCCCATTAAGGGTTCTGTGCTTA 1020
Db |||||||
Qy 961 AGAAGACTCATCTGCTAGTCTGTTGCTGACCTCTATGAGCCCATTAAGGGTTCTGTGCTTA 1020
Db |||||||
Qy 1021 GCATTAACAAATAAAGCTTTATAGTAAAGCCATGTATTAATTTTGTTCATGGAGG 1080
Db |||||||
Qy 1021 GCATTAACAAATAAAGCTTTATAGTAAAGCCATGTATTAATTTTGTTCATGGAGG 1080
Db |||||||
Qy 1081 GCCTTAAATTTGTGCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTCTAACTG 1140
Db |||||||
Qy 1081 GCCTTAAATTTGTGCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTCTAACTG 1140
Db |||||||
Qy 1141 CTTTTAGGAGATTAATTAATGATTAATAATTTAGTTTGGGGGAAATTAATTTGCGAAGA 1200
Db |||||||
Qy 1141 CTTTTAGGAGATTAATTAATGATTAATAATTTAGTTTGGGGGAAATTAATTTGCGAAGA 1200
Db |||||||
Qy 1201 GGATAATTTAATTTACGTCTCTGTTTATTCAGATAAAGAGAGAGACGCTACGCTGCATA 1260
Db |||||||
Qy 1201 GGATAATTTAATTTACGTCTCTGTTTATTCAGATAAAGAGAGAGACGCTACGCTGCATA 1260
Db |||||||
Qy 1261 TTCAAGAGTTGTACCTTAAACATTCGTGAAACATTTTCTAAGATTTTCAAAAGGAATAT 1320
Db |||||||
Qy 1261 TTCAAGAGTTGTACCTTAAACATTCGTGAAACATTTTCTAAGATTTTCAAAAGGAATAT 1320
Db |||||||
Qy 1321 GTGTAAATTTGAGAAATCATTAACCACTGCTCCTAAGTAAACAAACTGTTCTTAAATAA 1380
Db |||||||
Qy 1321 GTGTAAATTTGAGAAATCATTAACCACTGCTCCTAAGTAAACAAACTGTTCTTAAATAA 1380
Db |||||||
Qy 1381 AGTATTTAATGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTA 1434
Db |||||||
Qy 1381 AGTATTTAATGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTA 1434
Db |||||||

RESULT 2
ACC50424

ID ACC50424 standard; cDNA; 1434 BP.

XX ACC50424;

XX 12-JUN-2003 (first entry)

XX Human secreted protein coding sequence, SEQ ID 91.

XX Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
XX vulnery; antiinflammatory; nootropic; neuroprotective;
XX antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.
OS Homo sapiens.
XX WO200295010-A2.

RESULT 3
AAX37452

ID AAX37452 standard; cDNA; 1434 BP.

XX AC AAX37452;

XX DT 06-JUL-1999 (first entry)

XX DE Human secreted protein cDNA fragment containing gene 2.

XX KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
XX KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
XX KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
XX KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
XX KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
XX KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
XX KW arthritis; malignancy; digestive; endocrine; infection; ss.

XX OS Homo sapiens.

XX PN WO9918208-A1.

XX PD 15-APR-1999.

XX PF 01-OCT-1998; 98WO-US020775.

XX PR 02-OCT-1997; 97US-0060833P.

XX PR 02-OCT-1997; 97US-0060836P.

XX PR 02-OCT-1997; 97US-0060837P.

XX PR 02-OCT-1997; 97US-0060838P.

XX PR 02-OCT-1997; 97US-0060839P.

XX PR 02-OCT-1997; 97US-0060843P.

XX PR 02-OCT-1997; 97US-0060862P.

XX PR 02-OCT-1997; 97US-0060866P.

XX PR 02-OCT-1997; 97US-0060874P.

XX PR 02-OCT-1997; 97US-0060880P.

XX PR 02-OCT-1997; 97US-0060884P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;

XX PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;

XX PI Lafleur DW, Shi Y;

XX WPI; 1999-264022/22.

XX P-PSDB; AAY07853.

XX PT New isolated human genes and the secreted polypeptides they encode.

XX PS Claim 1a; Page 228; 368pp; English.

XX CC This invention describes novel isolated human genes and the secreted

XX CC proteins they encode. The products of the invention are useful for

XX CC preventing, treating or ameliorating medical conditions, e.g. by protein

XX CC or gene therapy. Also pathological conditions can be diagnosed by

XX CC determining the amount of the new polypeptides in a sample or by

XX CC determining the presence of mutations in the new polynucleotides.

XX CC Specific uses are described for each of the 101 polynucleotides, based on

XX CC which tissues they are most highly expressed in, and include developing

XX CC products for the diagnosis or treatment of cancer, tumours,

XX CC neurodegenerative disorders, developmental abnormalities and fetal

XX CC deficiencies, blood disorders, leukemias, diseases of the immune system,

XX CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,

XX CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate

XX CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,

XX CC transplant rejection, disorders involving osteoclasts such as

XX CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,

XX CC infections and AIDS. The human secreted proteins of the invention are

XX CC represented in AAY07852-Y07993 and the encoding nucleic acids are

XX CC represented in AAX37451-X37552

SQ Sequence 1434 BP; 480 A; 204 C; 250 G; 495 T; 0 U; 5 Other;

Query Match 96.4%; Score 1383; DB 2; Length 1434;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTCAATGTGCACCTCCATTCAATGTTGA 60

DB 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTCAATGTGCACCTCCATTCAATGTTGA 60

QY 61 TTTGGAGCTGACAGTATTTTGTGTAAGCAGAGATTTAAATTTTATATTGAAAGTCAGTGC 120

DB 61 TTTGGAGCTGACAGTATTTTGTGTAAGCAGAGATTTAAATTTTATATTGAAAGTCAGTGC 120

QY 121 AAAATTATGAATAGGATATCTAATAATCAAAAGTAAACAAAGTCAAAAGTCAGTGC 180

DB 121 AAAATTATGAATAGGATATCTAATAATCAAAAGTAAACAAAGTCAAAAGTCAGTGC 180

QY 181 CTAATAAAAAATCTCGGTTTCTTAAAAATTTTAAATTTTAAATTTTCTTGAATAGTCTTCT 240

DB 181 CTAATAAAAAATCTCGGTTTCTTAAAAATTTTAAATTTTAAATTTTCTTGAATAGTCTTCT 240

QY 241 TAGATTAACTCTCAGGATATGAGAAATCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300

DB 241 TAGATTAACTCTCAGGATATGAGAAATCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300

QY 301 CAAATTGCTTAAATGCAMGAGTGTATATATACAGAAATTTATCAGGCATTACCAAGTC 360

DB 301 CAAATTGCTTAAATGCAMGAGTGTATATATACAGAAATTTATCAGGCATTACCAAGTC 360

QY 361 TAGGCACATATAGGAAATGCAGCACTCAGAAATGTTTCAATGTAGTAGTTGATGTTGTA 420

DB 361 TAGGCACATATAGGAAATGCAGCACTCAGAAATGTTTCAATGTAGTAGTTGATGTTGTA 420

QY 421 AGGTAGGGAGCTTATTCAGACATATAGTATAGTATGTTTCTTAATGCTGTSTCAATTCGTC 480

DB 421 AGGTAGGGAGCTTATTCAGACATATAGTATAGTATGTTTCTTAATGCTGTSTCAATTCGTC 480

QY 481 CCTTTGCTACCTGCTACTTCCSANTATGGCAGCCCATTCAGTCTTCAGTCTTCTCTCT 540

DB 481 CCTTTGCTACCTGCTACTTCCSANTATGGCAGCCCATTCAGTCTTCTCTCTCTCTCTCT 540

QY 541 GGACACCTTATGCTCTGAAATCATGAGCGAGGCTGATTCATTCGTTGGTGGTAGA 600

DB 541 GGACACCTTATGCTCTGAAATCATGAGCGAGGCTGATTCATTCGTTGGTGGTAGA 600

QY 601 GCAGTATGTTTGTCTGACATTAAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT 660

DB 601 GCAGTATGTTTGTCTGACATTAAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT 660

QY 661 GTTTTTTACTTTTAAATAAGAAATATAACCTTTTAAAGTATTCACCTCTCTCCCCAGC 720

DB 661 GTTTTTTACTTTTAAATAAGAAATATAACCTTTTAAAGTATTCACCTCTCTCCCCAGC 720

QY 721 CTATCTCAAACCTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCACT 780

DB 721 CTATCTCAAACCTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCACT 780

QY 781 GCTCAGCTCAGGTAATCCGCCACCTCCACCTAGTATAGTATAGTATAGTATAGTATAGTATAGTAT 840

DB 781 GCTCAGCTCAGGTAATCCGCCACCTCCACCTAGTATAGTATAGTATAGTATAGTATAGTATAGTAT 840

QY 841 TTACTCTTCTTTCTGAAATCAGTTGTGAACCTGTTGCTTATGTTTCAAGTATTTAAGAAC 900

DB 841 TTACTCTTCTTTCTGAAATCAGTTGTGAACCTGTTGCTTATGTTTCAAGTATTTAAGAAC 900

QY 901 CTCMGTAATTCATTTTAAAAATCTGCTATCTGAGAAGCATTTGAATGAATCTTTAACA 960

DB 901 CTCMGTAATTCATTTTAAAAATCTGCTATCTGAGAAGCATTTGAATGAATCTTTAACA 960

QY 961 AGAAGACTCATCTAGTCTGTTCTGACTCTCTGACTCTCTGACTCTCTGACTCTCTGACTCTCTGACTCT 1020

DB 961 AGAAGACTCATCTAGTCTGTTCTGACTCTCTGACTCTCTGACTCTCTGACTCTCTGACTCTCTGACTCT 1020

PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 XX WPI; 2000-293155/25.
 XX Polynucleotide library comprising 1079 defined sequences, useful in the
 PT form of an array to detect cancer or susceptibility to cancer.
 XX
 XX Claim 1; Page 268; 502pp; English.
 XX
 XX The present invention describes a library of polynucleotides comprising
 CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
 CC one of the 1079 sequences; (2) a recombinant host cell containing (I);
 CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
 CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
 CC of detecting differentially expressed genes correlated with a cancerous
 CC state of a mammalian cell comprising detecting a gene product encoded by
 CC 65 of the 1079 sequences given in the specification. The polynucleotides
 CC are used to monitor patients having (or susceptible) to cancer to detect
 CC potentially malignant events at a molecular level before they are
 CC detectable at a gross morphological level. The polynucleotides are also
 CC useful for monitoring the efficacy of various therapies and preventive
 CC interventions. Polynucleotide probes based on the disclosed sequences are
 CC useful for chromosome mapping and detection of transcription levels. The
 CC 1079 polynucleotide sequences were derived from a human colon cancer cell
 CC line Kml2L4-A cDNA library
 XX
 XX Sequence 415 BP; 114 A; 73 C; 87 G; 141 T; 0 U; 0 Other;
 SQ
 Query Match 10.1%; Score 145; DB 3; Length 415;
 Best Local Similarity 100.0%; Pred. No. 4.7e-41;
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 324 GTGTAATATACAGAAATTTATCAGGCATTACCAAGTCTAGGCACATATAGGAATGCAGC 383
 DB 24 GTGTAATATACAGAAATTTATCAGGCATTACCAAGTCTAGGCACATATAGGAATGCAGC 83
 QY 384 ACTCAGATGTTTCAATGTAGTCTGATGCTTGTAGGTAGGGAGCTTATTCAGACA 443
 DB 84 ACTCAGATGTTTCAATGTAGTCTGATGCTTGTAGGTAGGGAGCTTATTCAGACA 143
 QY 444 TAGTAGATAGTCTTCTTAATGCTGT 468
 DB 144 TAGTAGATAGTCTTCTTAATGCTGT 168
 RESULT 6
 AAC10084
 ID AAC10084 standard; cDNA; 395 BP.
 AC AAC10084;
 XX
 XX 06-OCT-2000 (first entry)
 XX Human secreted protein 5' EST, SEQ ID NO: 14159.
 DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 KW Homo sapiens.
 OS
 XX EP1033401-A2.
 FN
 PD 06-SEP-2000.
 XX
 XX 21-FEB-2000; 2000EP-00200610.
 PF
 XX 26-FEB-1999; 99US-0122487P.
 PR
 XX (GEST) GENSET.
 PA
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI
 XX

DR WPI; 2000-500381/45.
 XX
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX
 XX Claim 1; SEQ ID NO 14159; 71pp + Sequence Listing; English.
 XX
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors
 XX
 XX Sequence 395 BP; 147 A; 41 C; 74 G; 125 T; 0 U; 8 Other;
 SQ
 Query Match 9.9%; Score 142; DB 3; Length 395;
 Best Local Similarity 100.0%; Pred. No. 5.2e-40;
 Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 58 TGATTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAATTTTATTTGAAGTCAG 117
 DB 1 TGATTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAATTTTATTTGAAGTCAG 60
 QY 118 TGCATAATTATGATAGATATATACTAATAATACAAAGTATAACAAAGTCAAGCAGT 177
 DB 61 TGCATAATTATGATAGATATATACTAATAATACAAAGTATAACAAAGTCAAGCAGT 120
 QY 178 GTTCTAATAAATAAATTCCTGGT 199
 DB 121 GTTCTAATAAATAAATTCCTGGT 142
 RESULT 7
 AAT25136
 ID AAT25136 standard; DNA; 336 BP.
 AC AAT25136;
 XX
 XX 22-OCT-1996 (first entry)
 XX Human gene signature HUMGS07294.
 DE
 XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 XX
 XX Homo sapiens.
 OS
 XX WO9514772-A1.
 FN
 PD 01-JUN-1995.
 XX
 XX 11-NOV-1994; 94WO-JP001916.
 PF
 XX 12-NOV-1993; 93JP-00355504.
 PR
 XX (MATS/) MATSUBARA K.
 PA (OKUBO/) OKUBO K.
 XX
 XX Matsubara K, Okubo K;
 PI
 XX WPI; 1995-206931/27.
 DR
 XX

PT Single-stranded DNA for identifying gene signatures - isolated from 3'-
 PT directed human cDNA library that reflects relative abundance of corresp.
 XX mRNA in specific human tissues.

PS Claim 1; Page 1777-78; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp. double
 CC -stranded DNA) which comprises one of the 7837 "GS" sequences given in
 CC AAT19001-726937 and which is able to hybridise to part of human genomic
 CC DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were
 CC obtained from 3'-directed cDNA libraries prepared from various human
 CC tissues; synthesis of cDNA was initiated from the 3'-end of mRNA by using
 CC poly(T) as the sole primer. Since the 3'-untranslated sequence is unique
 CC to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise
 CC with specific mRNAs. Each library is constructed so as to reflect
 CC accurately the relative abundance of different mRNAs in the particular
 CC tissue from which it was derived. The appearance frequency of a given GS
 CC in a cDNA library can be determined (esp. using primers and probes
 CC derived from the GS sequences) as a means of diagnosing abnormal cell
 CC function or for recognising different cell types

XX Sequence 336 BP; 95 A; 62 C; 62 G; 112 T; 0 U; 5 Other;

Query Match 8.9%; Score 127; DB 2; Length 336;

Best Local Similarity 100.0%; Pred. No. 8.7e-35;

Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 GTAATAAATCACTTCACTGCTCCACTCCAGTGGAATCGGCCACTCCCACTGACCTAGTA 822

Db 14 GTAAATAAATCACTTCACTGCTCCACTCCAGTGGAATCGGCCACTCCCACTGACCTAGTA 73

QY 823 GAATTTCTAATTTAATACCTTACCTTCTATTCTGAAATCAGTTGTAACCTGTCCTTAT 882

Db 74 GAATTTCTAATTTAATACCTTACCTTCTATTCTGAAATCAGTTGTAACCTGTCCTTAT 133

QY 883 GTTCAGA 889

Db 134 GTTCAGA 140

RESULT 8

AD881840/c

ID ADE81840 standard; cDNA; 493 BP.

XX ADE81840;

XX 29-JAN-2004 (first entry)

XX Arabidopsis thaliana expressed polynucleotide seq id 611.

XX genetically modified organism; transgenic organism; plant;
 KW inhibitor testing; activator testing; modifier testing; fungicide;
 KW insecticide; genetic function; genetic regulation; cellular metabolism;
 KW gene; ss.

XX Arabidopsis thaliana.

XX US2003115639-A1.

XX 19-JUN-2003.

XX 26-JAN-2001; 2001US-00770961.

XX 27-JAN-2000; 2000US-0178466P.

XX (GORL/) GORLACH J.

XX (ANY/) AN Y.

XX (HAMI/) HAMILTON C M.

XX (PRIC/) PRICE J L.

XX (RAIN/) RAINES T M.

XX (YUY/) YU Y.

XX (RAME/) RAMEKA J G.

XX (PAGE/) PAGE A.

PA (MATH/) MATHEW A V.
 PA (LEDF/) LEDFORD B L.
 PA (WOES/) WOESSNER J P.

PA (HAAS/) HAAS W D.

PA (GARC/) GARCIA C A.

PA (KRIC/) KRICKER M.

PA (SLAT/) SLATER T.

PA (DAVI/) DAVIS K R.

PA (ALLE/) ALLEN K.

PA (HOFF/) HOFFMAN N.

PA (HURB/) HURBAN P.

XX

PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;

PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;

PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;

PI Hurban P;

XX

DR WPI; 2003-810930/76.

XX

PT Novel Arabidopsis thaliana nucleic acids useful for generating

PT genetically modified transgenic organisms, for screening biologically

PT active agents such as fungicides, insecticides.

XX

PS Claim 1; SEQ ID NO 611; 44pp; English.

XX

CC The invention describes a nucleic acid (I) comprising a sequence capable

CC of hybridising under stringent conditions to any one of 999 fully defined

CC Arabidopsis thaliana sequences (I) as given in specification e.g., 360,

CC 1137, 455, 219, 472 nucleotides, etc, or its fragment. (I) is useful as a

CC hybridisation probe to complementary molecules in a cDNA library. (I) is

CC also useful for generating genetically modified and transgenic organisms,

CC usually plant cells and plants. A protein encoded by (I) is useful in

CC screening assays to determine the effect of candidate inhibitors,

CC activators or modifiers of the gene product. The protein is also useful

CC for screening biologically active agents e.g., fungicides and

CC insecticides. A genetically modified cell, comprising an exogenous

CC nucleic acid, where the nucleic acid comprises transcription regulatory

CC sequences operably linked to a sequence capable of hybridising under

CC stringent conditions to (I) is useful in the study of genetic function

CC and regulation, for alteration of the cellular metabolism and for

CC screening compounds that may affect the biological function of the gene

CC or gene product. This sequence represents an Arabidopsis thaliana

CC polynucleotide of the invention.

XX

SQ Sequence 493 BP; 146 A; 115 C; 70 G; 162 T; 0 U; 0 Other;

QY

Db

1390 TGATTTTAAAAA

52 TGATTTTAAAAA

Query Match 3.1%; Score 45; DB 9; Length 493;

Best Local Similarity 100.0%; Pred. No. 2.4e-06;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA

52 TGATTTTAAAAA

RESULT 9

AAH25190

ID AAH25190 standard; cDNA; 1231 BP.

XX

XX AAH25190;

XX

DT 22-AUG-2001 (first entry)

XX

DE Nucleotide sequence of a human transferrin.

XX

XX Human; transferrin; autoimmune disease; rheumatoid arthritis;

XX hyperproliferative disorder; neoplasm; cardiovascular disorder;

XX cardiac arrest; cerebrovascular disorder; cerebral ischemia;

XX angiogenesis; nervous system disorder; Alzheimer's disease; infection;

XX ocular disorder; corneal infection; wound healing;

XX epithelial cell proliferation; aging; organ transplant; ss.

XX Homo sapiens.

OS

```
XX FH Key Location/Qualifiers
XX CDS 3..989
XX FT /*tag= a
XX FT /product= "transferrin"
XX PN WO200146254-A1.
XX PD 28-JUN-2001.
XX PF 21-DEC-2000; 2000WO-US034769.
XX PR 23-DEC-1999; 99US-0171595P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Shi Y, Choi GH;
XX WPI; 2001-381910/40.
XX P-PSDB; AAB84371.
XX Isolated nucleic acid molecule encoding a human transferrin protein is
XX used in preventing, treating or ameliorating a medical condition.
XX Claim 1; Page 285; 290pp; English.
XX The present sequence encodes a human transferrin polypeptide. Transferrin
XX polypeptides and polynucleotides are used to prevent, treat or ameliorate
XX a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
XX dogs, chickens or sheep. Disorders which are diagnosed or treated include
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities
XX SQ Sequence 1231 BP; 408 A; 254 C; 305 G; 264 T; 0 U; 0 Other;

Query Match 3.1%; Score 45; DB 4; Length 1231;
Best Local Similarity 100.0%; Pred. No. 2e-06;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
DB 1147 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1191

RESULT 10
AAZ24411
ID AAZ24411 standard; cDNA; 1348 BP.
XX
XX AAZ24411;
XX
XX 14-FEB-2000 (first entry)
XX
XX Human bladder tumour cDNA library derived EST 23.
XX
XX Expressed sequence tag; human; bladder; tumour; cancer; cytostatic;
XX treatment; gene therapy; EST; ss.
XX
XX Homo sapiens.
XX
XX DE19818619-A1.
XX
XX 28-OCT-1999.
XX
XX 21-APR-1998; 98DE-01018619.

XX 21-APR-1998; 98DE-01018619.
XX (META-) METAGEN GES GENOMPORSCHUNG MBH.
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI; 1999-612028/53.
XX New nucleic acid sequences expressed in bladder tumor tissue, and derived
XX polypeptides, for treatment of bladder tumor and identification of
XX therapeutic agents.
XX Claim 3; Page 78; 132pp; German.
XX This invention describes novel polypeptide fragments (I) and the
XX polynucleotides (II) that encode them that are highly expressed in a
XX human bladder tumour and which have cytostatic activity. (II) are used
XX for recombinant expression of (I) and to isolate complete genes. (I) are
XX used to identify agents suitable for treatment of bladder cancer, to
XX directly treat this form of cancer (including expression from gene
XX therapy vectors) or are used in a preparation for cancer treatment. (I)
XX is also used for the generation of specific antibodies. (II) are
XX identified by assembling ESTs (expressed sequence tags) from a particular
XX tissue type before comparison of expression patterns. This allows a
XX significantly longer fragment of the gene to be revealed, and therefore
XX reduces the number of failures associated with the fact that ESTs from
XX different libraries may represent different parts of the same unknown
XX gene, distorting the estimated frequency of occurrence in a particular
XX tissue. AAZ3260-Z43309 represent expressed sequence tag (EST) fragments
XX isolated from a human bladder tumour cDNA library which encode the
XX proteins represented in AAY66143-Y66198
XX SQ Sequence 1348 BP; 326 A; 386 C; 349 G; 287 T; 0 U; 0 Other;

Query Match 3.1%; Score 45; DB 2; Length 1348;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
DB 1299 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1343

RESULT 11
ABL89915
ID ABL89915 standard; cDNA; 2552 BP.
XX
XX ABL89915;
XX
XX 24-MAY-2002 (first entry)
XX
XX Human polynucleotide SEQ ID NO 477.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US016450.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
```


XX WPI; 2002-122018/16.
DR P-PSDB; ABB89506.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX
XX Claim 4; SEQ ID NO 477; 2081pp + Sequence listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2552 BP; 807 A; 438 C; 512 G; 794 T; 0 U; 1 Other;

Query Match 3.1%; Score 45; DB 6; Length 2552;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
DB 2504 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2548

RESULT 12
ABV25066.
ID ABV25066 standard; cDNA; 3470 BP.
XX
AC ABV25066;
XX
DT 16-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 25057.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
(WILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX
XX Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX

XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 4868-4869; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 3470 BP; 1031 A; 694 C; 747 G; 938 T; 0 U; 60 Other;

Query Match 3.1%; Score 45; DB 5; Length 3470;
Best Local Similarity 100.0%; Pred. No. 1.6e-06;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
DB 3313 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3357

RESULT 13
AAA87997
ID AAA87997 standard; cDNA; 1730 BP.

XX
AC AAA87997;
XX
DT 08-DEC-2000 (first entry)
XX
DE Torenia hybrida flavone synthase encoding cDNA SEQ ID NO:3.
XX
KW Flavone; flavanone; flavone synthase; plant; flower colour;
KW bacterial resistance; nitrogen fixing bacteria; protection;
KW ultraviolet radiation; ss.
XX
OS Torenia hybrida.

PH Key Location/Qualifiers
FT CDS 57..1595
FT /*tag= a
FT /product= "flavone synthase"
XX
PN WO200044907-A1.
XX
PD 03-AUG-2000.

XX 28-JAN-2000; 2000WO-JP000490.
XX
XX 29-JAN-1999; 99JP-00022427.
PR 19-JUL-1999; 99JP-00205229.
XX
XX (SUNR) SUNTORY LTD.
XX
XX Mizutani M, Tanaka Y, Kusumi T, Ayabe S, Akashi T;
XX WPI; 2000-543394/49.
DR P-PSDB; AAB20573.

XX Gene encoding a flavone synthase for production of Antirrhinum and other
PT plants with modified flower colour.
XX
XX Claim 5; Page 34-38; 54pp; Japanese.
XX

QY 1392 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434

Search completed: July 20, 2004, 21:32:34
Job time : 589 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 21:11:14 ; Search time 128 Seconds
(without alignments)
6217.184 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaaactcttttttcg.....aaaaaaaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/prodata/2/ina/PCUTUS COMB.seq.*
- 6: /cgn2_6/prodata/2/ina/backfilees1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	3.1	1730	4	US-09-672-785-3
2	43	3.0	222	3	US-08-481-190-15
3	43	3.0	222	5	PCT-US93-00869-15
4	43	3.0	399	1	US-07-885-970A-13
5	43	3.0	399	1	US-08-298-687A-13
6	43	3.0	399	1	US-08-530-797-12
7	43	3.0	399	2	US-08-298-829-13
8	43	3.0	399	2	US-08-787-335-12
9	43	3.0	1114	4	US-09-152-060-41
10	43	3.0	1307	4	US-09-250-609-3
11	43	3.0	1307	4	US-09-250-611-3
12	43	3.0	1919	4	US-09-614-912-175
13	43	3.0	1976	4	US-09-920-759-10
14	43	3.0	2202	4	US-09-388-743-1
15	43	3.0	6200	4	US-09-439-923-1
16	43	3.0	6200	4	US-09-711-202A-1
17	43	3.0	6200	4	US-09-711-205A-1
18	42	2.9	144	1	US-08-702-344-26
19	42	2.9	258	4	US-09-621-976-15353
20	42	2.9	261	4	US-09-621-976-18330
21	42	2.9	276	4	US-09-621-976-18329
22	42	2.9	375	3	US-08-946-026-23
23	42	2.9	388	4	US-09-621-976-18573
24	42	2.9	396	4	US-09-640-173-10
25	42	2.9	396	4	US-09-640-173-42
26	42	2.9	396	4	US-09-713-550-10
27	42	2.9	396	4	US-09-713-550-42

```

28 42 2.9 769 4 US-08-567-882-5
29 42 2.9 911 2 US-08-924-759-9
30 42 2.9 911 3 US-09-248-335-9
31 42 2.9 991 3 US-08-924-747-25
32 42 2.9 991 3 US-09-247-373B-25
33 42 2.9 991 3 US-09-296-715-25
34 42 2.9 1035 2 US-08-580-545B-3
35 42 2.9 1035 3 US-09-262-653A-3
36 42 2.9 1035 3 US-08-867-484A-1
37 42 2.9 1035 4 US-09-834-859-1
38 42 2.9 1035 4 US-09-760-892-1
39 42 2.9 1035 4 US-09-834-656-1
40 42 2.9 1052 4 US-09-489-847-23
41 42 2.9 1174 2 US-08-872-437-1
42 42 2.9 1174 3 US-08-651-136C-11
43 42 2.9 1174 4 US-09-329-911A-11
44 42 2.9 1223 3 US-09-154-874-4
45 42 2.9 1223 4 US-08-931-668-4

```

ALIGNMENTS

```

RESULT 1
US-09-672-785-3
; Sequence 3, Application US/09672785
; Patent No. 6596927
; GENERAL INFORMATION:
; APPLICANT: Mizutani, Masako
; APPLICANT: Kasumi, Takaaki
; APPLICANT: Ayabe, Shin-ichi
; APPLICANT: Akashi, Tomoyoshi
; TITLE OF INVENTION: Genes Coding for Flavone Synthases
; FILE REFERENCE: 001560-383
; CURRENT APPLICATION NUMBER: US/09/672,785
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: PCT/JP00/04379
; PRIOR FILING DATE: 2000-01-30
; PRIOR APPLICATION NUMBER: PCT/JP00/00490
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: JP 11-205229
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: JP 11-22427
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1730
; TYPE: DNA
; ORGANISM: Torenia hybrida
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding a protein having an
; OTHER INFORMATION: activity to directly convert flavanone to flavone
US-09-672-785-3

```

```

Query Match 3.1%; Score 44; DB 4; Length 1730;
Best Local Similarity 100.0%; Pred. No. 7.4e-09;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1391 GATTTTAAAAA
DB 1680 GATTTTAAAAA

```

```

RESULT 2
US-08-481-190-15
; Sequence 15, Application US/08481190
; Patent No. 6160204
; GENERAL INFORMATION:
; APPLICANT: John C. Steffens
; TITLE OF INVENTION: Polyphenol Oxidase cDNA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:

```

Wed Jul 21 09:11:32 2004

ADDRESSEE: Yahwak & Associates
STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
FILING DATE: US/08/481,190
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 203,533
FILING DATE: 02-24-1994
ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: UA 816 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-08-481-190-15
Query Match 3.0%; Score 43; DB 3; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1392 ATTTTAA 1434
DB 172 ATTTTAA 214
RESULT 3
PCT-US93-00869-15
Sequence 15, Application PC/TUS9300869
GENERAL INFORMATION:
APPLICANT: John C. Steffens
TITLE OF INVENTION: Polyphenol Oxidase cDNAs: Cloning
TITLE OF INVENTION: and Applications
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Yahwak & Associates
STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: USA
ZIP: 06611
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
FILING DATE: PCT/US93/00869
APPLICATION NUMBER: 203,533
FILING DATE: 02-24-1994
ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: CRF D-1057
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951

TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
PCT-US93-00869-15
Query Match 3.0%; Score 43; DB 5; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1392 ATTTTAA 1434
DB 172 ATTTTAA 214
RESULT 4
US-07-885-970A-13
Sequence 13, Application US/07885970A
Patent No. 5495070
GENERAL INFORMATION:
APPLICANT: John, Maliyakal E.
TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nicholas J. Seay, Quarles & Brady
STREET: P.O. Box 2113, First Wisconsin Plaza
CITY: Madison
STATE: Wisconsin
COUNTRY: USA
ZIP: 53701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/885,970A
FILING DATE: 19920518
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/617,239
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/253,243
FILING DATE: 04-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 399 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Gossypium hirsutum
STRAIN: Coker 312
DEVELOPMENTAL STAGE: 10 day old fiber cells
TISSUE TYPE: fiber cells
IMMEDIATE SOURCE:
LIBRARY: CKFB10
CLONE: A9
US-07-885-970A-13

Query Match 3.0%; Score 43; DB 1; Length 399;
 Best Local Similarity 100.0%; Pred. No. 2.3e-08;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
 Db 345 ATTTTAAAAA 387

RESULT 5

US-08-298-687A-13
 ; Sequence 13, Application US/08298687A
 ; Patent No. 5521078
 ; GENERAL INFORMATION:
 ; APPLICANT: John, Maliyakal E.
 ; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
 ; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
 ; NUMBER OF SEQUENCES: 33
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
 ; STREET: P.O. Box 2113, First Wisconsin Plaza
 ; CITY: Madison
 ; STATE: Wisconsin
 ; COUNTRY: USA
 ; ZIP: 53701
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Microsoft Word
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/298,687A
 ; FILING DATE: 21-NOV-1990
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/617,239
 ; FILING DATE: 04-OCT-1988
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seay, Nicholas J.
 ; REGISTRATION NUMBER: 27,386
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (608) 283-2478
 ; TELEFAX: (608) 251-5139
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 399 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Gossypium hirsutum
 ; STRAIN: Coker 312
 ; DEVELOPMENTAL STAGE: 10 day old fiber cells
 ; TISSUE TYPE: fiber cells
 ; IMMEDIATE SOURCE:
 ; LIBRARY: CKFB10
 ; CLONE: A9
 ; US-08-298-687A-13

Query Match 3.0%; Score 43; DB 1; Length 399;
 Best Local Similarity 100.0%; Pred. No. 2.3e-08;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
 Db 345 ATTTTAAAAA 387

RESULT 6

US-08-530-797-12/C
 ; Sequence 12, Application US/08530797
 ; Patent No. 5597718
 ; GENERAL INFORMATION:
 ; APPLICANT: John, Maliyakal E.
 ; APPLICANT: Umbeck, Paul F.
 ; APPLICANT: Brill, Winston J.
 ; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
 ; TITLE OF INVENTION: FOR ALTERED FIBER
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Quarles and Brady
 ; STREET: P.O. Box 2113
 ; STREET: FIRST WISCONSIN PLAZA
 ; CITY: MADISON
 ; STATE: WISCONSIN
 ; COUNTRY: U.S.A.
 ; ZIP: 53701
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh
 ; SOFTWARE: Microsoft Word 4.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/530,797
 ; FILING DATE: 20-SEP-1995
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/617,239
 ; FILING DATE: 21-NOV-90
 ; APPLICATION NUMBER: US 07/253,243
 ; FILING DATE: 04-OCT-88
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Nicholas J. Seay
 ; REGISTRATION NUMBER: 27,386
 ; REFERENCE/DOCKET NUMBER: 1122990245
 ; INFORMATION FOR SEQ ID NO: 12:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 399 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; HYPOTHETICAL: no
 ; ANTI-SENSE: no
 ; ORIGINAL SOURCE:
 ; ORGANISM: Gossypium hirsutum
 ; STRAIN: Coker 312
 ; DEVELOPMENTAL STAGE: 10 day old fiber cells
 ; TISSUE TYPE: fiber cells
 ; IMMEDIATE SOURCE:
 ; LIBRARY: CKFB10
 ; CLONE: A9
 ; US-08-530-797-12

Query Match 3.0%; Score 43; DB 1; Length 399;
 Best Local Similarity 100.0%; Pred. No. 2.3e-08;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
 Db 55 ATTTTAAAAA 13

RESULT 7

US-08-298-829-13
 ; Sequence 13, Application US/08298829
 ; Patent No. 5620882
 ; GENERAL INFORMATION:
 ; APPLICANT: John, Maliyakal E.

Wed Jul 21 09:11:32 2004

;; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
;; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
;; NUMBER OF SEQUENCES: 33
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
;; STREET: P.O. Box 2113, First Wisconsin Plaza
;; CITY: Madison
;; STATE: Wisconsin
;; COUNTRY: USA
;; ZIP: 53701
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Microsoft Word
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/298,829
;; FILING DATE: 19-OCT-1994
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/885,970
;; FILING DATE: 18-MAY-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/617,239
;; FILING DATE: 21-NOV-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/253,243
;; FILING DATE: 04-OCT-1988
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seay, Nicholas J.
;; REGISTRATION NUMBER: 27,386
;; TELEPHONE: (608) 283-2478
;; TELEFAX: (608) 251-5139
;; INFORMATION FOR SEQ ID NO: 13:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 399 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Gossypium hirsutum
;; STRAIN: Coker 312
;; DEVELOPMENTAL STAGE: 10 day old fiber cells
;; TISSUE TYPE: fiber cells
;; IMMEDIATE SOURCE:
;; LIBRARY: CKFB10
;; CLONE: A9
;;
US-08-298-829-13

Query Match 3.0%; Score 43; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1392 ATTTTAAAAA 1434
Db 345 ATTTTAAAAA 387

RESULT 8
US-08-787-335-12/c
; Sequence 12, Application US/08/787335
; Patent No. 5981834
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; NUMBER OF SEQUENCES: 18

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Quarles and Brady
;; STREET: P.O. Box 2113
;; CITY: FIRST WISCONSIN PLAZA
;; STATE: WISCONSIN
;; COUNTRY: U.S.A.
;; ZIP: 53701
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
;; COMPUTER: Apple Macintosh
;; OPERATING SYSTEM: Macintosh
;; SOFTWARE: Microsoft Word 4.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/787,335
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/530,797
;; FILING DATE:
;; APPLICATION NUMBER: US 07/253,243
;; FILING DATE: 04-OCT-88
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Nicholas J. Seay
;; REGISTRATION NUMBER: 27,386
;; REFERENCE/DOCKET NUMBER: 1122990245
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 399 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA to mRNA
;; HYPOTHETICAL: no
;; ANTI-SENSE: no
;; ORIGINAL SOURCE:
;; ORGANISM: Gossypium hirsutum
;; STRAIN: Coker 312
;; DEVELOPMENTAL STAGE: 10 day old fiber cells
;; TISSUE TYPE: fiber cells
;; IMMEDIATE SOURCE:
;; LIBRARY: CKFB10
;; CLONE: A9
;;
US-08-787-335-12

Query Match 3.0%; Score 43; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1392 ATTTTAAAAA 1434
Db 55 ATTTTAAAAA 13

RESULT 9
US-09-152-060-41
; Sequence 41, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PZ003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100

```
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,357
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,189
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/057,765
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/068,368
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-152-060-41

Query Match      3.0%; Score 43; DB 4; Length 1114;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 1029 ATTTTAAAAA 1071

RESULT 10
US-09-250-609-3
; Sequence 3, Application US/09250609A
; Patent No. 6458943
; GENERAL INFORMATION:
; APPLICANT: Byrnie, Jennifer A.
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210002
; CURRENT APPLICATION NUMBER: US/09/250,609A
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1307
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (192)..(803)
US-09-250-609-3

Query Match      3.0%; Score 43; DB 4; Length 1307;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 1263 ATTTTAAAAA 1305

RESULT 11
US-09-250-611-3
; Sequence 3, Application US/09250611
; Patent No. 6528283
; GENERAL INFORMATION:
; APPLICANT: Byrnie, Jennifer A.
; APPLICANT: Basset, Paul
; TITLE OF INVENTION: Members of the D52 Gene Family
; FILE REFERENCE: 1383.0210001
; CURRENT APPLICATION NUMBER: US/09/250,611
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1307
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (192)..(803)
US-09-250-611-3

Query Match      3.0%; Score 43; DB 4; Length 1307;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 1263 ATTTTAAAAA 1305

RESULT 12
US-09-614-912-175
; Sequence 175, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 175
; LENGTH: 1919
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-614-912-175

Query Match      3.0%; Score 43; DB 4; Length 1919;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 1876 ATTTTAAAAA 1918

RESULT 13
US-09-920-759-10
; Sequence 10, Application US/09920759
; Patent No. 6537811
; GENERAL INFORMATION:
; APPLICANT: Brenda F. Baker
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF SAP-1 EXPRESSION
; FILE REFERENCE: RTS-0267
```

```
; CURRENT APPLICATION NUMBER: US/09/920,759
; CURRENT FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 91
; SEQ ID NO 10
; LENGTH: 1976
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (150)...(1367)
US-09-920-759-10

Query Match
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 3.0%; Score 43; DB 4; Length 1976;
1921 ATTTTAAAAA 100.0%; Pred. No. 1.6e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1392 ATTTTAAAAA 3.0%; Score 43; DB 4; Length 6200;
1921 ATTTTAAAAA 100.0%; Pred. No. 1.6e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 3.0%; Score 43; DB 4; Length 6200;
1921 ATTTTAAAAA 100.0%; Pred. No. 1.6e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1392 ATTTTAAAAA 3.0%; Score 43; DB 4; Length 6200;
1921 ATTTTAAAAA 100.0%; Pred. No. 1.6e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: July 21, 2004, 00:04:40
Job time : 129 secs
```

```
US-09-388-743-1
; Sequence 1, Application US/09388743
; Patent No. 6423886
; GENERAL INFORMATION:
; APPLICANT: Singletary, George
; APPLICANT: Zhou, Lan
; TITLE OF INVENTION: No. 6423886el Starch Synthase Polynucleotides and Their
; TITLE OF INVENTION: Use in the Production of New Starches
; FILE REFERENCE: 1144
; CURRENT APPLICATION NUMBER: US/09/388,743
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2202
; TYPE: DNA
; ORGANISM: Curcuma zedoaria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (130)...(1974)
US-09-388-743-1

Query Match
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 3.0%; Score 43; DB 4; Length 2202;
1921 ATTTTAAAAA 100.0%; Pred. No. 1.8e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1392 ATTTTAAAAA 3.0%; Score 43; DB 4; Length 2202;
1921 ATTTTAAAAA 100.0%; Pred. No. 1.8e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: July 21, 2004, 00:04:40
Job time : 129 secs
```

```
US-09-439-923-1
; Sequence 1, Application US/09439923
; Patent No. 6426208
; GENERAL INFORMATION:
; APPLICANT: Emil D. Kakkis
; APPLICANT: Becky Tanamachi
; TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods
; TITLE OF INVENTION: for Producing and Purifying the Same and Methods for
; TITLE OF INVENTION: Treating Diseases Caused by Deficiencies Thereof
; FILE REFERENCE: 08000051US00
; CURRENT APPLICATION NUMBER: US/09/439,923
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6200
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 23:04:30 ; Search time 677 Seconds
(without alignments)

10333.750 Million cell updates/sec

Title: US-10-799-747-12

Perfect score: 1434

Sequence: 1 cattaactctttttatcg.....aaaaaaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3191023 seqs, 2439312756 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6382046

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Published Applications NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
 - 12: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
 - 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
 - 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
 - 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
 - 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
 - 17: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
 - 18: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
 - 19: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
 - 20: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1434	100.0	1434	15	US-10-195-730-12
2	345	24.1	448	9	US-09-880-107-1385
3	47	3.3	8281	13	US-10-342-887-309
4	47	3.3	8281	13	US-10-172-118-309
5	45	3.1	493	10	US-09-770-961-611
6	45	3.1	1231	9	US-09-891-126-3
7	45	3.1	1231	9	US-10-266-745-3
8	45	3.1	2552	16	US-10-264-237-477
9	44	3.1	1730	16	US-10-411-115-3
10	43	3.0	193	10	US-09-814-353-18984
11	43	3.0	240	13	US-10-424-599-44257
12	43	3.0	282	10	US-09-814-353-17457
13	43	3.0	294	10	US-09-814-353-5106
14	43	3.0	294	10	US-09-814-353-11398

c 15	43	3.0	306	16	US-10-131-827-8716	Sequence 8716, Ap
c 16	43	3.0	337	9	US-09-770-791-889	Sequence 889, App
c 17	43	3.0	338	13	US-10-424-599-53632	Sequence 53632, A
c 18	43	3.0	341	17	US-10-021-323-451	Sequence 451, App
c 19	43	3.0	342	17	US-10-021-323-9783	Sequence 9783, Ap
c 20	43	3.0	344	13	US-10-424-599-35199	Sequence 35199, A
c 21	43	3.0	361	13	US-10-424-599-12247	Sequence 12247, A
c 22	43	3.0	374	13	US-10-424-599-50183	Sequence 50183, A
c 23	43	3.0	382	13	US-10-424-599-67598	Sequence 67598, A
c 24	43	3.0	382	13	US-10-424-599-82042	Sequence 82042, A
c 25	43	3.0	390	17	US-10-437-963-95670	Sequence 95670, A
c 26	43	3.0	398	17	US-10-021-323-7863	Sequence 7863, Ap
c 27	43	3.0	402	10	US-09-918-995-16715	Sequence 5673, Ap
c 28	43	3.0	403	13	US-10-424-599-56715	Sequence 16715, A
c 29	43	3.0	408	17	US-10-437-963-82810	Sequence 82810, A
c 30	43	3.0	415	13	US-10-424-599-5972	Sequence 5972, Ap
c 31	43	3.0	432	13	US-10-424-599-107176	Sequence 107176, A
c 32	43	3.0	433	13	US-10-424-599-122494	Sequence 122494, A
c 33	43	3.0	455	13	US-10-424-599-105065	Sequence 105065, A
c 34	43	3.0	457	17	US-10-437-963-113203	Sequence 113203, A
c 35	43	3.0	459	10	US-09-814-353-17965	Sequence 50630, A
c 36	43	3.0	480	16	US-10-131-827-8192	Sequence 17965, A
c 37	43	3.0	489	17	US-10-021-323-17555	Sequence 17555, A
c 38	43	3.0	492	17	US-10-437-963-29142	Sequence 29142, A
c 39	43	3.0	502	13	US-10-424-599-33446	Sequence 33446, A
c 40	43	3.0	510	17	US-10-437-963-26905	Sequence 26905, A
c 41	43	3.0	522	17	US-10-021-323-1498	Sequence 1498, Ap
c 42	43	3.0	526	17	US-10-021-323-14764	Sequence 14764, A
c 43	43	3.0	528	13	US-10-424-599-46393	Sequence 46393, A
c 44	43	3.0	528	16	US-10-131-827-8186	Sequence 8186, Ap
c 45	43	3.0	528	16	US-10-131-827-8186	Sequence 8186, Ap

ALIGNMENTS

RESULT 1

US-10-195-730-12
; Sequence 12, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; PRIOR FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-195-730-12

Query Match 100.0%; Score 1434; DB 15; Length 1434;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CATTAACTCTTTTATCGGAATAGTAGATATTTCAATGTCTCATTCATTTGA	60
Db	1	CATTAACTCTTTTATCGGAATAGTAGATATTTCAATGTCTCATTCATTTGA	60
Qy	61	TTTGGAGCTGACAGTATTTTGTCTACGACAGATTTTATTTTATTTGAAGTCAGTGC	120
Db	61	TTTGGAGCTGACAGTATTTTGTCTACGACAGATTTTATTTTATTTGAAGTCAGTGC	120
Qy	121	AAAATTATGATAGATATATACTAATAAATAACAAAGTCAAGACAGTGT	180


```
; Sequence 309, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 309
; LENGTH: 8281
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-309

Query Match          3.3%; Score 47; DB 13; Length 8281;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AATGATTTTAAAAA 1434
Db 8219 AATGATTTTAAAAA 8265

RESULT 4
US-10-172-118-309
; Sequence 309, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 309
; LENGTH: 8281
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AL133623
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-309

Query Match          3.3%; Score 47; DB 13; Length 8281;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AATGATTTTAAAAA 1434
Db 8219 AATGATTTTAAAAA 8265
```

```
RESULT 5
US-09-770-961-611/c
; Sequence 611, Application US/09770961
; Publication No. US20030115639A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2026 (PARA-015PRV)
; CURRENT APPLICATION NUMBER: US/09/770,961
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,466
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 611
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-961-611

Query Match          3.1%; Score 45; DB 10; Length 493;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434
Db 52 TGATTTTAAAAA 8

RESULT 6
US-09-891-126-3
; Sequence 3, Application US/09891126
; Patent No. US20020072596A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Transferrin Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT035PI
; CURRENT APPLICATION NUMBER: US/09/891,126
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: PCT/US00/34769
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/171,595
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-126-3

Query Match          3.1%; Score 45; DB 9; Length 1231;
```

Best Local Similarity 100.0%; Pred. No. 2.3e-10; Mismatches 0; Indels 0; Gaps 0;

Matches 45; Conservative 0;

QY 1390 TGATTTTAAAAA 1434

Db 1147 TGATTTTAAAAA 1191

RESULT 7

US-10-266-745-3

Sequence 3, Application US/10266745

Publication No. US20030149256A1

GENERAL INFORMATION:

APPLICANT: Ruben et al.

TITLE OF INVENTION: Transferrin Polynucleotides, Polypeptides, and Antibodies

FILE REFERENCE: PT035P1

CURRENT APPLICATION NUMBER: US/10/266,745

CURRENT FILING DATE: 2002-10-09

PRIOR APPLICATION NUMBER: US/09/891,126

PRIOR FILING DATE: 2001-06-26

PRIOR APPLICATION NUMBER: PCT/US00/34769

PRIOR FILING DATE: 2000-12-21

PRIOR APPLICATION NUMBER: 60/171,595

PRIOR FILING DATE: 1999-12-23

NUMBER OF SEQ ID NOS: 7

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 3

LENGTH: 1231

TYPE: DNA

ORGANISM: Homo sapiens

US-10-266-745-3

Query Match 3.1%; Score 45; DB 15; Length 1231;

Best Local Similarity 100.0%; Pred. No. 2.3e-10;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434

Db 1147 TGATTTTAAAAA 1191

RESULT 8

US-10-264-237-477

Sequence 477, Application US/10264237

Publication No. US20040009491A1

GENERAL INFORMATION:

APPLICANT: Birse et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PA131P1

CURRENT APPLICATION NUMBER: US/10/264,237

CURRENT FILING DATE: 2002-10-04

PRIOR APPLICATION NUMBER: PCT/US01/16450

PRIOR FILING DATE: 2001-05-18

PRIOR APPLICATION NUMBER: US 60/205,515

PRIOR FILING DATE: 2000-05-19

NUMBER OF SEQ ID NOS: 2876

SOFTWARE: Patentin Ver. 3.1

SEQ ID NO 477

LENGTH: 2552

TYPE: DNA

ORGANISM: Homo sapiens

US-10-264-237-477

Query Match 3.1%; Score 45; DB 16; Length 2552;

Best Local Similarity 100.0%; Pred. No. 2.5e-10;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434

Db 2504 TGATTTTAAAAA 2548

RESULT 9

US-10-411-115-3

Sequence 3, Application US/10411115

Publication No. US20040003431A1

GENERAL INFORMATION:

APPLICANT: Mizutani, Masako

APPLICANT: Kasumi, Takaaki

APPLICANT: Ayabe, Shin-ichi

APPLICANT: Akashi, Tomoyoshi

TITLE OF INVENTION: Genes Coding for Flavone Synthases

FILE REFERENCE: 001560-383

CURRENT APPLICATION NUMBER: US/10/411,115

CURRENT FILING DATE: 2003-04-11

PRIOR APPLICATION NUMBER: US/09/672,785

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: PCT/JP00/04379

PRIOR FILING DATE: 2000-01-30

PRIOR APPLICATION NUMBER: PCT/JP00/00490

PRIOR FILING DATE: 1999-01-28

PRIOR APPLICATION NUMBER: JP 11-205229

PRIOR FILING DATE: 1999-01-19

PRIOR APPLICATION NUMBER: JP 11-22427

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 8

SOFTWARE: Patentin version 3.0

SEQ ID NO 3

LENGTH: 1730

TYPE: DNA

ORGANISM: Torenia hybrida

FEATURE:

OTHER INFORMATION: Nucleotide sequence encoding a protein having an activity to directly convert flavanone to flavone

US-10-411-115-3

Query Match 3.1%; Score 44; DB 16; Length 1730;

Best Local Similarity 100.0%; Pred. No. 7e-10;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1391 GATTTTAAAAA 1434

Db 1680 GATTTTAAAAA 1723

RESULT 10

US-09-814-353-18984/c

Sequence 18984, Application US/09814353

Publication No. US20030165831A1

GENERAL INFORMATION:

APPLICANT: Thompson, Pamela

APPLICANT: Lillie, James

TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF OVARIAN CANCER

FILE REFERENCE: MRI-006B

CURRENT APPLICATION NUMBER: US/09/814,353

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: US 60/191,031

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: US 60/207,124

PRIOR FILING DATE: 2000-05-25

PRIOR APPLICATION NUMBER: US 60/211,940

PRIOR FILING DATE: 2000-06-15

PRIOR APPLICATION NUMBER: US 60/216,820

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: US 60/220,661

PRIOR FILING DATE: 2000-07-25

PRIOR APPLICATION NUMBER: US 60/257,672

NUMBER OF SEQ ID NOS: 22037

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 18984

LENGTH: 193

TYPE: DNA

Best Local Similarity 100.0%; Pred. No. 2.3e-10; Mismatches 0; Indels 0; Gaps 0;

Matches 45; Conservative 0;

QY 1390 TGATTTTAAAAA 1434

Db 1147 TGATTTTAAAAA 1191

RESULT 7

US-10-266-745-3

Sequence 3, Application US/10266745

Publication No. US20030149256A1

GENERAL INFORMATION:

APPLICANT: Ruben et al.

TITLE OF INVENTION: Transferrin Polynucleotides, Polypeptides, and Antibodies

FILE REFERENCE: PT035P1

CURRENT APPLICATION NUMBER: US/10/266,745

CURRENT FILING DATE: 2002-10-09

PRIOR APPLICATION NUMBER: US/09/891,126

PRIOR FILING DATE: 2001-06-26

PRIOR APPLICATION NUMBER: PCT/US00/34769

PRIOR FILING DATE: 2000-12-21

PRIOR APPLICATION NUMBER: 60/171,595

PRIOR FILING DATE: 1999-12-23

NUMBER OF SEQ ID NOS: 7

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 3

LENGTH: 1231

TYPE: DNA

ORGANISM: Homo sapiens

US-10-266-745-3

Query Match 3.1%; Score 45; DB 15; Length 1231;

Best Local Similarity 100.0%; Pred. No. 2.3e-10;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434

Db 1147 TGATTTTAAAAA 1191

RESULT 8

US-10-264-237-477

Sequence 477, Application US/10264237

Publication No. US20040009491A1

GENERAL INFORMATION:

APPLICANT: Birse et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PA131P1

CURRENT APPLICATION NUMBER: US/10/264,237

CURRENT FILING DATE: 2002-10-04

PRIOR APPLICATION NUMBER: PCT/US01/16450

PRIOR FILING DATE: 2001-05-18

PRIOR APPLICATION NUMBER: US 60/205,515

PRIOR FILING DATE: 2000-05-19

NUMBER OF SEQ ID NOS: 2876

SOFTWARE: Patentin Ver. 3.1

SEQ ID NO 477

LENGTH: 2552

TYPE: DNA

ORGANISM: Homo sapiens

US-10-264-237-477

Query Match 3.1%; Score 45; DB 16; Length 2552;

Best Local Similarity 100.0%; Pred. No. 2.5e-10;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434

Db 2504 TGATTTTAAAAA 2548

RESULT 9

US-10-411-115-3

Sequence 3, Application US/10411115

Publication No. US20040003431A1

GENERAL INFORMATION:

APPLICANT: Mizutani, Masako

APPLICANT: Kasumi, Takaaki

APPLICANT: Ayabe, Shin-ichi

APPLICANT: Akashi, Tomoyoshi

TITLE OF INVENTION: Genes Coding for Flavone Synthases

FILE REFERENCE: 001560-383

CURRENT APPLICATION NUMBER: US/10/411,115

CURRENT FILING DATE: 2003-04-11

PRIOR APPLICATION NUMBER: US/09/672,785

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: PCT/JP00/04379

PRIOR FILING DATE: 2000-01-30

PRIOR APPLICATION NUMBER: PCT/JP00/00490

PRIOR FILING DATE: 1999-01-28

PRIOR APPLICATION NUMBER: JP 11-205229

PRIOR FILING DATE: 1999-01-19

PRIOR APPLICATION NUMBER: JP 11-22427

PRIOR FILING DATE: 1999-01-29

NUMBER OF SEQ ID NOS: 8

SOFTWARE: Patentin version 3.0

SEQ ID NO 3

LENGTH: 1730

TYPE: DNA

ORGANISM: Torenia hybrida

FEATURE:

OTHER INFORMATION: Nucleotide sequence encoding a protein having an activity to directly convert flavanone to flavone

US-10-411-115-3

Query Match 3.1%; Score 44; DB 16; Length 1730;

Best Local Similarity 100.0%; Pred. No. 7e-10;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1391 GATTTTAAAAA 1434

Db 1680 GATTTTAAAAA 1723

RESULT 10

US-09-814-353-18984/c

Sequence 18984, Application US/09814353

Publication No. US20030165831A1

GENERAL INFORMATION:

APPLICANT: Thompson, Pamela

APPLICANT: Lillie, James

TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF OVARIAN CANCER

FILE REFERENCE: MRI-006B

CURRENT APPLICATION NUMBER: US/09/814,353

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: US 60/191,031

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: US 60/207,124

PRIOR FILING DATE: 2000-05-25

PRIOR APPLICATION NUMBER: US 60/211,940

PRIOR FILING DATE: 2000-06-15

PRIOR APPLICATION NUMBER: US 60/216,820

PRIOR FILING DATE: 2000-07-07

PRIOR APPLICATION NUMBER: US 60/220,661

PRIOR FILING DATE: 2000-07-25

PRIOR APPLICATION NUMBER: US 60/257,672

NUMBER OF SEQ ID NOS: 22037

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 18984

LENGTH: 193

TYPE: DNA

```
; ORGANISM: Homo sapiens
US-09-814-353-18984

Query Match          3.0%; Score 43; DB 10; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 170 ATTTTAAAAA 128

RESULT 11
US-10-424-599-44257
; Sequence 44257, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 44257
; LENGTH: 240
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_139963C.1
US-10-424-599-44257

Query Match          3.0%; Score 43; DB 13; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 27 ATTTTAAAAA 69

RESULT 12
US-09-814-353-17457/c
; Sequence 17457, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17457
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 126, 127, 128, 142, 143, 144, 145, 146, 147, 148, 149, 162,
; LOCATION: 174, 176, 177, 178, 181, 199, 206, 213, 223, 227, 228, 232,
; LOCATION: 235, 255
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-17457/c

Query Match          3.0%; Score 43; DB 10; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 124 ATTTTAAAAA 82

RESULT 14
US-09-814-353-11398/c
; Sequence 11398, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17457
```

```
; LENGTH: 282
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-17457

Query Match          3.0%; Score 43; DB 10; Length 282;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 217 ATTTTAAAAA 175

RESULT 13
US-09-814-353-5106/c
; Sequence 5106, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5106
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 126, 127, 128, 142, 143, 144, 145, 146, 147, 148, 149, 162,
; LOCATION: 174, 176, 177, 178, 181, 199, 206, 213, 223, 227, 228, 232,
; LOCATION: 235, 255
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-5106

Query Match          3.0%; Score 43; DB 10; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 124 ATTTTAAAAA 82

RESULT 14
US-09-814-353-11398/c
; Sequence 11398, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5106
```

Search completed: July 21, 2004, 01:37:12
Job time : 679 secs

```

; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11398
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 126, 127, 128, 142, 143, 144, 145, 146, 147, 148, 149, 162,
; LOCATION: 174, 176, 177, 178, 181, 199, 206, 213, 223, 227, 228, 232,
; LOCATION: 235, 255
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-11398

Query Match 3.0%; Score 43; DB 10; Length 294;
Best Local Similarity 100.0%; Pred.No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 124 ATTTTAAAAA 82

RESULT 15
US-10-131-827-8716/c
; Sequence 8716, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
; APPLICANT: Wohlgemuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: LV, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 56661200120
; CURRENT APPLICATION NUMBER: US/10/131,827
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8716
; LENGTH: 306
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-8716

Query Match 3.0%; Score 43; DB 16; Length 306;
Best Local Similarity 100.0%; Pred.No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 1434
Db 67 ATTTTAAAAA 25
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 21:04:06 ; Search time 3468 Seconds
(without alignments)
12347.857 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaaactctttttatcgg.....aaaaaaaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST*

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estnu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_htc:**
- 9: gb_estl:**
- 10: gb_est2:**
- 11: gb_htc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estcom:**
- 17: em_gss_hum:**
- 18: em_gss_inv:**
- 19: em_gss_pln:**
- 20: em_gss_vrt:**
- 21: em_gss_fun:**
- 22: em_gss_mam:**
- 23: em_gss_mus:**
- 24: em_gss_pro:**
- 25: em_gss_rod:**
- 26: em_gss_pbg:**
- 27: em_gss_vrl:**
- 28: gb_gssl:**
- 29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query	ID	Description
C 1	448	31.2	814	14	CB853053
C 2	444	31.0	444	9	AA136080
C 3	421	29.4	551	12	BI259796
C 4	414	28.9	556	10	AW500190

C 5	401	28.0	588	10	BF431622
C 6	400	27.9	438	10	AW304923
C 7	395	27.5	395	9	AA928811
C 8	364	25.4	400	9	AW002842
C 9	346	24.1	1201	13	EX441923
C 10	345	24.1	448	9	AA491000
C 11	336	23.4	785	12	BG484396
C 12	313	21.8	941	13	EX452125
C 13	303	21.1	419	10	AW189910
C 14	286	19.9	568	9	AL035942
C 15	284	19.8	557	14	CB132625
C 16	276	19.2	280	9	AA679646
C 17	254	17.7	493	9	AW026264
C 18	240	16.7	394	9	AA747222
C 19	227	15.8	621	9	AL035941
C 20	227	15.8	688	12	BM728856
C 21	203	14.2	575	12	BM722991
C 22	200	13.9	253	9	AA953459
C 23	181	12.6	503	10	AW889139
C 24	170	11.9	765	12	BI914473
C 25	149	10.4	957	12	BG403642
C 26	105	7.3	1037	13	EX413464
C 27	104	7.3	338	12	BI025209
C 28	82	5.7	634	13	EX506904
C 29	77	5.4	240	10	BF910533
C 30	60	4.2	1117	12	BM466406
C 31	48	3.3	419	10	AW189910
C 32	47	3.3	229	12	BG736035
C 33	47	3.3	229	12	BG737212
C 34	47	3.3	427	10	BE031403
C 35	46	3.2	231	14	CF350700
C 36	46	3.2	281	12	BI679838
C 37	46	3.2	386	10	AW733394
C 38	46	3.2	423	12	BI774735
C 39	46	3.2	439	14	CF602204
C 40	46	3.2	488	10	BE723220
C 41	46	3.2	868	14	CK160202
C 42	45	3.1	102	9	AI447690
C 43	45	3.1	116	14	CD520351
C 44	45	3.1	134	9	AI499325
C 45	45	3.1	162	13	BQ189083

ALIGNMENTS

RESULT 1	CB853053/c	814 bp	mRNA	linear	EST 22-APR-2003
LOCUS	UI-CF-FNO-afg-e-20-0-UI-s1	UI-CF-FNO	Homo sapiens	cdna clone	
DEFINITION	UI-CF-FNO-afg-e-20-0-UI 3', mRNA sequence.				
ACCESSION	CB853053				
VERSION	CB853053.1	GI:30048171			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 814)				
TITLE	Bonaldi,M.F., Lennon,G. and Soares,M.B.				
JOURNAL	Normalization and subtraction: two approaches to facilitate gene discovery				
MEDLINE	Genome Res. 6 (9), 791-806 (1996)				
PUBMED	9704477				
COMMENT	889548				
	Contact: McCray, PB				
	McCray Lab				
	University of Iowa				
	2024 University of Iowa Med Labs, Iowa City, IA 52242, USA				
	Tel: 319 356 4866				
	Fax: 319 356 7171				
	Email: paul-mccray@uiowa.edu				
	Tissue Procurement: Dr. M. J. Welsh, University of Iowa				


```

QY 1023 ATTACAAATAGGTTTATAGTAAAGCAATGTATTATTTTTTTCATGGAGGC 1082
Db 384 ATTACAAATAGGTTTATAGTAAAGCAATGTATTATTTTTTTCATGGAGGC 325
QY 1083 TTTAAATTTGCTCTTTTTCATATTTTATTCATATTCATATTTGTTGTAAGTCT 1142
Db 324 TTTAAATTTGCTCTTTTTCATATTTTATTCATATTCATATTTGTTGTAAGTCT 265
QY 1143 TTTTGGAGAGATAATTATATGTATATAATTTAGTTTGGGGGAATTAATTTGTCAGAGAG 1202
Db 264 TTTTGGAGAGATAATTATATGTATATAATTTAGTTTGGGGGAATTAATTTGTCAGAGAG 205
QY 1203 ATAAATTTTACGCTCTTTTTCATATTTTATTCAGATAAAGAGAGAGACTACGCTGCATAT 1262
Db 204 ATAAATTTTACGCTCTTTTTCATATTTTATTCAGATAAAGAGAGAGACTACGCTGCATAT 145
QY 1263 CAAGAGTTGTACCTTAAACATTTGTTGTAACATTTTCTAAGATTTTCAAAGAGATATGT 1322
Db 144 CAAGAGTTGTACCTTAAACATTTGTTGTAACATTTTCTAAGATTTTCAAAGAGATATGT 85
QY 1323 GTAAATTTGAGAATCATTAACCACTGCTCACTTGTGTAACCACTGCTTAAATAAAG 1382
Db 84 GTAAATTTGAGAATCATTAACCACTGCTCACTTGTGTAACCACTGCTTAAATAAAG 25
QY 1383 TATTTAATGATTTTAAAAA 1406
Db 24 TATTTAATGATTTTAAAAA 1

RESULT 3
B1259796 551 bp mRNA linear EST 17-JUL-2001
LOCUS 602971440F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5110904 5',
DEFINITION mRNA sequence.
ACCESSION B1259796
VERSION B1259796.1 GI:14817472
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 551)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM1268 row: p column: 09
High quality sequence stop: 536.
Location/Qualifiers
1..551
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5110904"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_lib="NIH_MGC_12"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
ORIGIN
Query Match 29.4%; Score 421; DB 12; Length 551;

```

```

Best Local Similarity 100.0%; Pred. No. 0;
Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1000 CCCATAAGGGTCTGTGCTTAGCATTAACAAATTAAGGTTTATAGTAAAGCAATGTAT 1059
Db 120 CCCATAAGGGTCTGTGCTTAGCATTAACAAATTAAGGTTTATAGTAAAGCAATGTAT 179
QY 1060 TAATTTTTTTTGGCATGGAGGCTTTAAATTTTGCTCTCTTTTTCATATTTTATTCATAT 1119
Db 180 TAATTTTTTTTGGCATGGAGGCTTTAAATTTTGCTCTCTTTTTCATATTTTATTCATAT 239
QY 1120 TCAATTTATGTTTGTAACTGCTTTTACGGAGATAATTAATTTACGTCTCTGTTTATCAGATAA 1239
Db 240 TCAATTTATGTTTGTAACTGCTTTTACGGAGATAATTAATTTACGTCTCTGTTTATCAGATAA 359
QY 1180 GGGGGAATTAATTTGTCAAAGAGAGATAATTAATTTACGTCTCTGTTTATCAGATAA 1239
Db 300 GGGGGAATTAATTTGTCAAAGAGAGATAATTAATTTACGTCTCTGTTTATCAGATAA 359
QY 1240 GAGAGAAGACTACGCTGTCATATTCAAGAGTTTGTACCTTAACATTTGTTGTAAGCAATTTTTC 1299
Db 360 GAGAGAAGACTACGCTGTCATATTCAAGAGTTTGTACCTTAACATTTGTTGTAAGCAATTTTTC 419
QY 1300 TAAGATTTTCAAAGAGATAATGTTGTAATTTGAGAATCATACCACTGCTCTTAACCTTGGTGAACATTTTTC 1359
Db 420 TAAGATTTTCAAAGAGATAATGTTGTAATTTGAGAATCATACCACTGCTCTTAACCTTGGT 479
QY 1360 AAACAACTGTTCTTAAATTAAGTATTTAAATGATTTTAAAAA 1419
Db 480 AAACAACTGTTCTTAAATTAAGTATTTAAATGATTTTAAAAA 1419
QY 1420 A 1420
Db 540 A 540

RESULT 4
AW500190/c 556 bp mRNA linear EST 01-MAR-2000
LOCUS UI-HF-BNO-akc-g-07-0-UI.rl NIH_MGC_50 Homo sapiens cDNA clone
DEFINITION IMAGE:3076789 5', mRNA sequence.
ACCESSION AW500190
VERSION AW500190.1 GI:7112570
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 556)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
cDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
The following repetitive elements were found in this cDNA
sequence:
8-48, >AT_rich#Low_complexity
Seq primer: M13 Forward.
Location/Qualifiers
1..556
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3076789"
/tissue_type="lymph"
FEATURES
source

```


Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyt not found
 Seq primer: -40UP from Gibco
 High quality sequence stop: 410.

FEATURES

source
 1. 438
 /location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2826421"
 /tissue types="three pooled meningiomas"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP Brn53"
 /notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies."

ORIGIN

Query Match 27.9%; Score 400; DB 10; Length 438;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1000 CCATAGGGTTCGTGCTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAT 1059
 Db
 QY 1060 TAAATTTTTTTCATGGAGGGCTTTAAATTTGTGCTCTTTTCATATTTTTCATAT 1119
 Db
 QY 1120 TCAATTTATGGTTTGAACGCTTTTGGAGGATAATATATGTTATATAATTTAGTTTG 1179
 Db
 QY 1180 GGGGGATAATTTGCAAGAGGATAATTTAATTTAGCTCTCTCTGTTATTCAGATAAA 1239
 Db
 QY 1240 GAGAGAAGACTACGCTGCATTTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTTC 1299
 Db
 QY 1300 TAAGATTTTCAAAGGAATATGTTAAATTTGAGAAATCATACCACTGCTTAACCTTGGT 1359
 Db
 QY 1360 AAACAACTGTTCTTAAATAAGTATTTAATGATTTTAA 1399
 Db
 QY 40 AAACAACCTGTTCTTAAATAAGTATTTAATGATTTTAA 1

RESULT 7

AA92811/c
 LOCUS
 DEFINITION
 IMAGE:1623533 3', mRNA sequence.
 AA92811
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 395)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

JOURNAL
 COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 500 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 375.

FEATURES

source
 1. 395
 /location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1623533"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"
 /clone_lib="Soares total fetus Nb2HF8 9w"
 /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'] Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaído."

ORIGIN

Query Match 27.5%; Score 395; DB 9; Length 395;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1011 TCTGTGCTTAGCATTAACAAATAAGGTTTATAGTAAAGCCCAATGTATTAATTTTTTT 1070
 Db
 QY 1071 TGCATCGAGGGCTTTAAATTTGTGCTCTTTTCATATTTTATCATATTCAATTTATGG 1130
 Db
 QY 1131 TTGTAACTGCTTTTATAGGAGATAATATATGTTATAAATAGTTTGGGGGGAATAAT 1190
 Db
 QY 1191 TGTGCAAGAGGATAATTTAATTTACGTCTTCTGTATTTCAGAAATAAGAGAGAGAGACT 1250
 Db
 QY 1251 ACCTGTCATATTCAGAGTTGTACCTTAACATTTGGTGAACATTTTCTAGATTTCAT 1310
 Db
 QY 1311 AAAGGAATATGTGTAATTTGAGAAATCATACCACTGCTTAACCTGCTTAACCTGTTAAACAACTGT 1370
 Db
 QY 1371 TCTTAAATAAGATTTAATGATTTTAAAAA 1405
 Db
 QY 35 TCTTAAATAAGATTTAATGATTTTAAAAA 1

RESULT 8

AW002842/c
 LOCUS
 DEFINITION
 IMAGE:1623533 3', mRNA sequence.
 AW002842
 VERSION
 KEYWORDS
 SOURCE
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 395)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.


```

QY 923 ATCTGCTATTCTGAGAGCAATTGAATGAATCTTAAACAAGAGACTCATCTCTAGTGT 982
|||||
Db 562 ATCTGCTATTCTGAGAGCAATTGAATGAATCTTAAACAAGAGACTCATCTCTAGTGT 621
|||||
QY 983 TGCTGACTCTCTATGAGCCC 1001
|||||
Db 622 TGCTGACTCTCTATGAGCCC 640

RESULT 10
AA491000/c
LOCUS AA491000.1 linear EST 18-AUG-1997
DEFINITION aa5911.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824612 3',
mRNA sequence.
ACCESSION AA491000
VERSION AA491000.1 GI:2220173
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 448)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 436.
FEATURES
Location/Qualifiers
1..448
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:824612"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI CGAP GCB1"
/notes="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CEER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATGAAGTGGAGCGCGCCCTCATTTTTTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN
Query Match 24.1%; Score 345; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1000 CCATAGGGTCTGCTAGCATTAAACAATAAGGTTATAGTAAAGCCAAATGTAT 1059
|||||
Db 414 CCATAGGGTCTGCTAGCATTAAACAATAAGGTTATAGTAAAGCCAAATGTAT 355
|||||
QY 1060 TANTTTTTTTCATGGAGGGCTTAABATTTTGCTCTCTTTTCATATTATTTCATAT 1119
|||||

```

```

Db 354 TAAATTTTTTTTGCATGAGAGGCTTTAAATTTTGCTCTTTTTCATATTTTTCATAT 295
|||||
QY 1120 TCAATTTTATGGTTTGTAACTGCTTTTATAGGAGATAAATTATATGTTTATAAATTAGTTTG 1179
|||||
Db 294 TCAATTTATGGTTTGTAACTGCTTTTATAGGAGATAAATTATATGTTTATAAATTAGTTTG 235
|||||
QY 1180 GGGGGAATAAATTGTGCAAGAGAGATAAATTAAATTACGTGCTTCGTATTACAGATAAA 1239
|||||
Db 234 GGGGGAATAAATTGTGCAAGAGAGATAAATTAAATTACGTGCTTCGTATTACAGATAAA 175
|||||
QY 1240 GAGAGAAGACTACGCTGCATATTCAAGAGTCTTACCTTAACATTCGTGTAACATTTTTC 1299
|||||
Db 174 GAGAGAAGACTACGCTGCATATTCAAGAGTCTTACCTTAACATTCGTGTAACATTTTTC 115
|||||
QY 1300 TAAAGATTTTCAAAAGGAATATGTAAATTGAGAATCATACCA 1344
|||||
Db 114 TAAAGATTTTCAAAAGGAATATGTAAATTGAGAATCATACCA 70
|||||

RESULT 11
BG484396
LOCUS BG484396
DEFINITION 602505037F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618473 5',
mRNA sequence.
ACCESSION BG484396
VERSION BG484396.1 GI:13416675
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 785)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1376 row: j column: 10
High quality sequence start: 18
High quality sequence stop: 751.
FEATURES
Location/Qualifiers
1..785
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4618473"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/notes="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattagggc); Site 2: SfiI (ggccattagggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGGGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
Kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
ORIGIN
Query Match 23.4%; Score 336; DB 12; Length 785;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 324 GTGGTAATATACAGATTTATCAGGCAATTCACAGCTAGGCACATATAGGAAATGCAGC 383
|||||

```

Db 139 GTGGTAATACAGAAATTTATCAGGCAATACCAAGCTAGGCACATATAGGAATGCAGC 198
QY 384 ACTCAGAAATGGTTTCAATGTAGTAGTGTGCTTGTAGGTAGGGAGCTTATTCAGACA 443
Db 199 ACTCAGAAATGGTTTCAATGTAGTAGTGTGCTTGTAGGTAGGGAGCTTATTCAGACA 258
QY 444 TAGTAGATATTTCTCTCAATGCTGTGTCAATGTGCTGCTTGTGCTACCTGTTCCSC 503
Db 259 TAGTAGATATTTCTCTCAATGCTGTGTCAATGTGCTGCTTGTGCTACCTGTTCCSC 318
QY 504 ATTATGCGACCCCAATTCAGCTTTGAGTTTTCTCTCTGACACCTTATGCTCTGAATCA 563
Db 319 ATTATGCGACCCCAATTCAGCTTTGAGTTTTCTCTCTGACACCTTATGCTCTGAATCA 378
QY 564 TGAAGGAGGCTGATTCATTTGGTGTGATTTGGTGTAGAAAGCAGTAGTGTCTGACATTAA 623
Db 379 TGAAGGAGGCTGATTCATTTGGTGTGATTTGGTGTAGAAAGCAGTAGTGTCTGACATTAA 438
QY 624 GATGTAGGTTATAGATAGGTTTAAAGCTTTTAAAGTGTATGTTTTTATCTTTAAATAAGAA 683
Db 439 GATGTAGGTTATAGATAGGTTTAAAGCTTTTAAAGTGTATGTTTTTATCTTTAAATAAGAA 498
QY 684 ATATAACCTTTTAAAGCTTATCCACCTCTCCCGAGCTATCTCAAACTGGTGGAAATATA 743
Db 499 ATATAACCTTTTAAAGCTTATCCACCTCTCCCGAGCTATCTCAAACTGGTGGAAATATA 558
QY 744 TGGAGAGATCTTGAAGA 761
Db 559 TGGAGAGATCTTGAAGA 576

RESULT 12
BX452125 941 bp mRNA linear EST 22-MAY-2003
LOCUS BX452125 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS0DF023Y002 5-PRIME, mRNA sequence.
ACCESSION BX452125
VERSION BX452125.1 GI:31034315
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 941)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2532.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG018ZG08_CS01695_1&cluster=2532.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0BAG018ZG08_CS01695_1.
Location/Qualifiers
1. .941
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF023Y002"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

FEATURES
source

ORIGIN

Query Match 21.8%; Score 313; DB 13; Length 941;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1122 AATTTATGTTTCTCAATGCTGTGTCAATGTGCTGCTTGTGCTACCTGTTCCSC 1181
Db 23 AATTTATGTTTCTCAATGCTGTGTCAATGTGCTGCTTGTGCTACCTGTTCCSC 82
QY 1182 GCGAATAAATGTGCAAGAGGATAATTTAATTTACGTGCTTCTGTTATTTCAGATAAAGA 1241
Db 83 GCGAATAAATGTGCAAGAGGATAATTTAATTTACGTGCTTCTGTTATTTCAGATAAAGA 142
QY 1242 GAGAAGACTACGCTGCATATTTCAAGAGTTGTACCTTAAACATTGGTGAAACATTTTCTA 1301
Db 143 GAGAAGACTACGCTGCATATTTCAAGAGTTGTACCTTAAACATTGGTGAAACATTTTCTA 202
QY 1302 AGATTTTCAAAAGGAATATCTGTAAATTCAGAAATCATACCACTGCTTAACCTTGTAA 1361
Db 203 AGATTTTCAAAAGGAATATCTGTAAATTCAGAAATCATACCACTGCTTAACCTTGTAA 262
QY 1362 ACAACCTGTTCTTAAATAAAGTATTTAATGATTTTAAATAAAGTATTTAATGATTTTAA 1421
Db 263 ACAACCTGTTCTTAAATAAAGTATTTAATGATTTTAAATAAAGTATTTAATGATTTTAA 322
QY 1422 AAAAAA 1434
Db 323 AAAAAA 335

RESULT 13
AW189910/c

LOCUS AW189910 419 bp mRNA linear EST 22-NOV-1999
DEFINITION x110a04.x1 NCI CGAP Ut4 Homo sapiens cDNA clone IMAGE:2675790 3',
similar to contains element LTR5 repetitive element ;, mRNA
sequence.
ACCESSION AW189910
VERSION AW189910.1 GI:6464390
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 419)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA library preparation: Life Technologies, Inc.
cDNA library arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .419
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2675790"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP Ut4"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.48 kb. Life Technologies catalog #:

```
ORIGIN
Query Match 11542-016"
Best Local Similarity 21.1%; Score 303; DB 10; Length 419;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1099 TTTTTCATATTATTCATATTCATATTTATGTTTGTAACTGCTTTTATAGGAGAGATAATT 1158
DB 303 TTTTTCATATTATTCATATTCATATTTATGTTTGTAACTGCTTTTATAGGAGAGATAATT 244

QY 1159 ATATGTTATAAATTAGTTTGGGGGAATAATTGTCGAAAGAGATAATTAAATTACGT 1218
DB 243 ATATGTTATAAATTAGTTTGGGGGAATAATTGTCGAAAGAGATAATTAAATTACGT 184

QY 1219 GCTTCTGTTATTCAGATAAAGAGAGAGACACCTGTCATATTCAGAGATTGTACCTTA 1278
DB 183 GCTTCTGTTATTCAGATAAAGAGAGAGACACCTGTCATATTCAGAGATTGTACCTTA 124

QY 1279 ACATTGGTGAAACATTTTCTAAGATTTTCAAAGGAATATGTGTAAATTGAGAAATCA 1338
DB 123 ACATTGGTGAAACATTTTCTAAGATTTTCAAAGGAATATGTGTAAATTGAGAAATCA 64

QY 1339 TAACCACTGCTCTAACTGGTAAACAACTGTTCTTAAATAAAGTATTTAAATGATTTTAA 1398
DB 63 TAACCACTGCTCTAACTGGTAAACAACTGTTCTTAAATAAAGTATTTAAATGATTTTAA 4

QY 1399 AAA 1401
DB 3 AAA 1

RESULT 14
AL035942/c
LOCUS
DEFINITION 568 bp mRNA linear EST 04-SEP-2003
ACCESSION DKF2p564B1622_s1 564 (synonym: hfbr2) Homo sapiens cDNA clone
VERSION DKF2p564B1622 3', mRNA sequence.
KEYWORDS AL035942
SOURCE AL035942.1 GI:5405576
ORGANISM Homo sapiens (human)
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Wambutt, R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
JOURNAL EST (Wambutt, et al.)
COMMENT Unpublished (1999)
CONTACT: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
rl sequence also available.
This clone (DKF2p564B1622) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1. .568
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKF2p564B1622"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="X1-2blue"
/clone_lib="564 (synonym: hfbr2)"
/note="Vector: pAMP1; Site_1: NotI; Site_2: SalI"

ORIGIN
Query Match 19.9%; Score 286; DB 9; Length 568;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 336; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1000 CCCATAAGGGTTCTGCTTAGCATTAAACAAAATAAGGTTTATAGGTAAGCAATGAT 1059
DB 413 CCCATAAGGGTTCTGCTTAGCATTAAACAAAATAAGGTTTATAGGTAAGCAATGAT 354

QY 1060 TAAATTTTTTTCATGAGAGGGCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATAT 1119
DB 353 TAAATTTTTTTCATGAGAGGGCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATAT 294

QY 1120 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATATATATAGTTTGTG 1179
DB 293 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATATATATAGTTTGTG 234

QY 1180 GGGGGAATAAATTTGTCAAAGAGGATAAATTTAATTTACGTGCTTCTGTTATTCAGAAATAA 1239
DB 233 GGGGGAATAAATTTGTCAAAGAGGATAAATTTAATTTACGTGCTTCTGTTATTCAGAAATAA 174

QY 1240 GAGAGAAGACTACGCTGCATATTCAGAGATTGTACCTTAAACATTTGTTGAAACATTTTTC 1299
DB 173 GAGAGAAGACTACGCTGCATATTCAGAGATTGTACCTTAAACATTTGTTGAAACATTTTTC 114

QY 1300 TAAAGATTTTCAAAGGAATATGTGTAAATTTGAGAAAT 1336
DB 113 TAAAGATTTTCAAAGGAATATGTGTAAATTTGAGAAAT 77

RESULT 15
CB132625
LOCUS
DEFINITION 557 bp mRNA linear EST 29-JAN-2003
K-EST0183064 L14Choick0 Homo sapiens cDNA clone L14Choick0-2-D11
5', mRNA sequence.
ACCESSION CB132625
VERSION CB132625.1 GI:28097898
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,
JOURNAL Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and
COMMENT Kim, Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
CONTACT: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 2 row: D column: 11
High quality sequence stop: 557.
FEATURES
source
1. .557
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L14Choick0-2-D11"
/sex="M"
/cell_line="Choi-CK"
/lab_host="Top10f"
/clone_lib="L14Choick0"
/note="Organ: Liver; Vector: pT73-Pac; Site_1: EcoRI;
Site_2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN
```

Query Match 19.8%; Score 284; DB 14; Length 557;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 554; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 674 AAAATAAGAAATATAACCTTTTAAGCTATTCACCTCTCTCCCGAGCTATCTCAAACTG 733
Db |||||||
1 AAAATAAGAAATATAACCTTTTAAGCTATTCACCTCTCTCCCGAGCTATCTCAAACTG 60

QY 734 GTGGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCACTGCTCCACTCCAGG 793
Db |||||||

QY 61 GTGGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCACTGCTCCACTCCAGG 120
Db |||||||

QY 794 TGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTTAATACTTACCTTCTATTT 853
Db |||||||

QY 121 TGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTTAATACTTACCTTCTATTT 180
Db |||||||

QY 854 CTGAATCAGTTGTGAAGCTGTGCTTATGTTGAGAGTTTAAGAACTCTGAGTGAATTC 913
Db |||||||

QY 181 CTGAATCAGTTGTGAAGCTGTGCTTATGTTGAGAGTTTAAGAACTCTGAGTGAATTC 240
Db |||||||

QY 914 TTTTAAAAATCTGCTATTTCTGAGAAGCTTGAATGAATTTCTTAACTTAACAAAT 1033
Db |||||||

QY 241 TTTTAAAAATCTGCTATTTCTGAGAAGCTTGAATGAATTTCTTAACTTAACAAAT 359
Db |||||||

QY 974 GTAGCTGTTTGTGCTGACTCCCTATGAG -CCCATAGGGTTCTGTGCTTAGCAATTAACAAAT 1033
Db |||||||

QY 1034 AAGGTTTATAGGTAAAGCCCAATGATTAATTTTTCATGAGAGGCTTTAAATTTG 1093
Db |||||||

QY 360 AAGGTTTATAGGTAAAGCCCAATGATTAATTTTTCATGAGAGGCTTTAAATTTG 419
Db |||||||

QY 1094 TGCTCTTTTTCATATTTTATTCATATTCATTTATGTTTGTAACTGCTTTTAGGGAGA 1153
Db |||||||

QY 420 TGCTCTTTTTCATATTTTATTCATATTCATTTATGTTTGTAACTGCTTTTAGGGAGA 479
Db |||||||

QY 1154 TAATTATATGTTAATAATTAAGTTTGGGGGAATAATTTGCAAGAGGATATTTAAT 1213
Db |||||||

QY 480 TAATTATATGTTAATAATTAAGTTTGGGGGAATAATTTGCAAGAGGATATTTAAT 539
Db |||||||

QY 1214 TAGGTGCTTCTGTTATTC 1231
Db |||||||

QY 540 TAGGTGCTTCTGTTATTC 557
Db |||||||

Search completed: July 21, 2004, 00:02:22
Job time : 3471 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 16:59:02 ; Search time 52 Seconds
(without alignments)
108.672 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Genesepi1980s.*

2: Genesepi1990s.*

3: Genesepi2000s.*

4: Genesepi2001s.*

5: Genesepi2002s.*

6: Genesepi2003s.*

7: Genesepi2003bs.*

8: Genesepi2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	99.0	19	6 ADA40708	Ada40708 Human sec
2	98	99.0	19	6 ABR47713	Abr47713 Human sec
3	98	99.0	20	2 AAY07853	Aay07853 Human sec
4	51	51.5	355	2 AAW29179	Aaw29179 Rat CC ch
5	51	51.5	355	7 ADD45360	Add45360 Rat Prote
6	49	49.5	2235	7 ADC56729	Adc56729 Thale cre
7	49	49.5	2235	7 ADC56731	Adc56731 Thale cre
8	48	48.5	422	2 AAR67688	Aar67688 S. pombe
9	48	48.5	422	2 AAR67687	Aar67687 S. pombe
10	47	47.5	26	6 ABP55266	Abp55266 Human mus
11	47	47.5	148	4 ABG14555	Abg14555 Novel hum
12	47	47.5	317	2 AAR48689	Aar48689 G-protein
13	47	47.5	317	2 AAW02661	Aaw02661 G-protein
14	47	47.5	456	7 ADD48620	Add48620 Human pro
15	47	47.5	466	1 AAP96203	Aap96203 Human mus
16	47	47.5	466	4 AAB20483	Aab20483 Cholinerg
17	47	47.5	466	4 ABB56361	Abb56361 Non-endog
18	47	47.5	466	6 ABB99723	Abb99723 Amino aci
19	47	47.5	466	6 ABP81853	Abp81853 Human mus
20	47	47.5	466	7 ADE54491	Ade54491 Rat Prote
21	47	47.5	466	7 ADD48618	Add48618 Rat Prote
22	47	47.5	466	7 ADE54493	Ade54493 Human pro
23	47	47.5	496	5 ABG31502	Abg31502 Muscarini
24	47	47.5	500	5 ABG31297	Abg31297 Hydrophob
25	46.5	47.0	32	4 AAM99595	Aam99595 Human exc

26	46.5	47.0	32	4 AAM42418	Aam42418 Human kid
27	46	46.5	333	2 AAR48691	Aar48691 G-protein
28	46	46.5	333	2 AAW02663	Aaw02663 G-protein
29	46	46.5	345	7 ADC86023	Adc86023 Human GPC
30	46	46.5	422	2 AAR88131	Aar88131 Aureobasi
31	46	46.5	422	2 AAR88130	Aar88130 Aureobasi
32	46	46.5	478	5 AAU97292	Aau97292 Reference
33	46	46.5	478	5 AAU97293	Aau97293 Polymorph
34	46	46.5	478	6 ABB99725	Abb99725 Amino aci
35	46	46.5	478	6 ABP81854	Abp81854 Human mus
36	46	46.5	479	1 AAP92606	Aap92606 Human mus
37	46	46.5	479	4 ABB56363	Abb56363 Non-endog
38	45	45.5	28	2 AAW40000	Aaw40000 Peptide e
39	45	45.5	209	3 AAB58414	Aab58414 Lung canc
40	45	45.5	295	4 AAG80106	Aag80106 Human CCR
41	45	45.5	355	2 AAR52749	Aar52749 C-C chemo
42	45	45.5	355	2 AAW25751	Aaw25751 Human MIP
43	45	45.5	355	2 AAW26588	Aaw26588 Human MIP
44	45	45.5	355	3 AAB20571	Aab20571 Human CC-
45	45	45.5	355	6 ABU09082	Abu09082 Human che

ALIGNMENTS

RESULT 1

ADA40708	ID	ADA40708	standard; protein; 19 AA.
XX	AC	ADA40708;	
XX	DT	20-NOV-2003	(first entry)
XX	DE	Human secreted protein.	
XX	KW	Human; secreted protein; cancer; hyperproliferative disorder;	
KW	rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;		
KW	anaemia; allergic reaction; asthma; cardiovascular disorder;		
KW	wound healing; cytostatic; immunosuppressive; neurotropic; neuroprotective;		
KW	antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;		
XX	vulnery; cardiant; gene therapy.		
OS	Homo sapiens.		
XX	EN	WO2002102993-A2.	
XX	PD	27-DEC-2002.	
XX	PF	19-MAR-2002; 2002WO-US008123.	
XX	PR	21-MAR-2001; 2001US-0277340P.	
XX	PR	19-JUL-2001; 2001US-0306171P.	
XX	PR	13-NOV-2001; 2001US-0331287P.	
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Rosen CA, Ruben SM;	
XX	DR	WPI; 2003-175238/17.	
XX	PT	New human secreted proteins and nucleic acid molecules, useful for	
XX	PT	preparing a diagnostic or pharmaceutical composition for diagnosing,	
XX	PT	preventing or treating cancer or other hyperproliferative disorder,	
XX	PT	asthma, allergies or AIDS.	
XX	PS	Claim 1; SEQ ID NO 1090; 3205pp; English.	
XX	CC	The invention relates to novel genes ADA39629-ADA40565 and proteins	
XX	CC	ADA40566-ADA41501 for human secreted proteins, useful for preventing,	
XX	CC	treating or ameliorating medical conditions e.g. by protein or gene	
XX	CC	therapy. The polypeptides, nucleic acid molecules, antibodies or their	
XX	CC	fragments, and agonists or antagonists that bind to the polypeptide are	
XX	CC	useful for preparing a diagnostic or pharmaceutical composition for	

diagnosing or treating cancer or other hyperproliferative disorder. The polypeptides and nucleic acid molecules are also useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating cancer or other hyperproliferative disorders including neoplasms, autoimmune disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic anaemia), haematopoietic or haematological disorders (e.g. anaemia, thrombocytopenia), allergic reactions including asthma or eczema, inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory bowel disease or Crohn's disease), neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular disorders (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial, fungal or viral infections including HIV/AIDS), or wound healing and disorders of epithelial cell proliferation. The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping, as molecular weight markers, or as hybridization or diagnostic probes. The polypeptides and antibodies are useful for providing immunological probes for differential identification of the tissues immunohistochemistry assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Query Match 99.0%; Score 98; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAHSLVSLFLWTPYALKS 19
| | | | | | | | | | | | | | | | | | |
Db 1 MAHSLVSLFLWTPYALKS 19

RESULT 2
ABR47713
ID ABR47713 standard; protein; 19 AA.
XX
AC ABR47713;
XX
DT 12-JUN-2003 (first entry)
XX
DE Human secreted protein, SEQ ID 604.
XX
KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
KW vulnary; antiinflammatory; nootropic; neuroprotective;
KW antiparkinsonian; gene therapy; human; cardiovascular disorder.
XX
OS Homo sapiens.
XX
FN WO200295010-A2.
XX
PD 28-NOV-2002.
XX
PE 19-MAR-2002; 2002WO-US009785.
XX
FR 21-MAR-2001; 2001US-0277340P.
FR 19-JUL-2001; 2001US-0306171P.
FR 13-NOV-2001; 2001US-0331287P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2003-129429/12.
XX
PT Novel human secreted proteins, useful for detecting, preventing,
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
PT disorders such as arrhythmia.
XX
PS Claim 13; SEQ ID NO 604; 1881pp; English.
XX
XX The present invention relates to novel human secreted proteins (ABR47633-

ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins and their coding sequences are useful for the preparation of a diagnostic or pharmaceutical composition for diagnosing or treating a cardiovascular disorder (e.g. arrhythmia, tachycardia, cardiac arrest, coronary arteriosclerosis and myocardial ischaemia), neural disorders, immune system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, pulmonary disorders, renal disorders, CC proliferative disorders and/or cancerous diseases and conditions, for wound healing and epithelial cell proliferation, to treat inflammation or infection, for treating thrombosis and arteriosclerosis, for treating or preventing neural damage which occurs in neuronal disorders or neurodegenerative conditions such as Alzheimer's disease and Parkinson's disease, to enhance bone and periodontal regeneration and aid in tissue transplants or bone grafts, to prevent skin aging or hair loss, to stimulate growth and differentiation of haematopoietic cells and bone marrow cells when used in combination with other cytokines, to maintain organs before transplantation or for supporting cell culture of primary tissues, to increase or decrease differentiation or proliferation of embryonic stem cells, or to modulate mammalian characteristics or metabolism. Note: The sequence data for this patent was published in electronic format and is available from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 19 AA;

Query Match 99.0%; Score 98; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAHSLVSLFLWTPYALKS 19
| | | | | | | | | | | | | | | | | | |
Db 1 MAHSLVSLFLWTPYALKS 19

RESULT 3
AAV07853
ID AAV07853 standard; protein; 20 AA.
XX
AC AAV07853;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein fragment encoded from gene 2.

Human; secreted protein; treatment; prevention; protein therapy; AIDS; gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder; developmental abnormality; fetal deficiency; blood disorder; leukemia; immune system disease; autoimmune disease; hepatic disease; lymphoma; renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia; cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder; pulmonary disorder; transplant rejection; osteoclast; osteoporosis; arthritis; malignancy; digestive; endocrine; infection.

Homo sapiens.
XX
OS WO9918208-A1.
XX
PD 15-APR-1999.
XX
PE 01-OCT-1998; 98WO-US020775.
XX
PR 02-OCT-1997; 97US-0060833P.
PR 02-OCT-1997; 97US-0060836P.
PR 02-OCT-1997; 97US-0060837P.
PR 02-OCT-1997; 97US-0060838P.
PR 02-OCT-1997; 97US-0060839P.
PR 02-OCT-1997; 97US-0060843P.
PR 02-OCT-1997; 97US-0060862P.
PR 02-OCT-1997; 97US-0060866P.
PR 02-OCT-1997; 97US-0060874P.
PR 02-OCT-1997; 97US-0060880P.
PR 02-OCT-1997; 97US-0060884P.

PA (HUMA-) HUMAN GENOME SCI INC.
 XX Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
 PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
 PI Lafleur DW, Shi Y;
 XX WPI; 1999-264022/22.
 DR N-PSDB; AAX37452.
 XX
 PT New isolated human genes and the secreted polypeptides they encode.
 XX
 PS Claim 1b; Page 276; 368pp; English.
 XX
 CC This invention describes novel isolated human genes and the secreted
 CC proteins they encode. The products of the invention are useful for
 CC preventing, treating or ameliorating medical conditions, e.g. by protein
 CC or gene therapy. Also pathological conditions can be diagnosed by
 CC determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 101 polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, tumours,
 CC neurodegenerative disorders, developmental abnormalities and fetal
 CC deficiencies, blood disorders, leukemias, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
 CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
 CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
 CC transplant rejection, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
 CC infections and AIDS. The human secreted proteins of the invention are
 CC represented in AAY07852-Y07993 and the encoding nucleic acids are
 CC represented in AAX37451-X37552
 XX
 SQ Sequence 20 AA;
 XX
 Query Match 99.0%; Score 98; DB 2; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.9e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 MAHSVLSFLLWTPYALKS 19
 DB 1 MAHSVLSFLLWTPYALKS 19
 XX
 RESULT 4
 AAW29179
 ID AAW29179 standard; protein; 355 AA.
 XX
 AC AAW29179;
 XX
 DT 19-DEC-1997 (first entry)
 XX
 DE Rat CC chemokine receptor.
 XX
 KW rat; CC chemokine receptor; screen; binding; ligand.
 XX
 OS Rattus rattus.
 XX
 PN JP09227599-A.
 XX
 PD 02-SEP-1997.
 XX
 PF 22-FEB-1996; 96JP-00035192.
 XX
 PR 22-FEB-1996; 96JP-00035192.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 DR WPI; 1997-486426/45.
 DR N-PSDB; AAT86839.
 XX
 PT CC chemokine receptor protein - useful to screen for novel binding
 compounds.

XX
 PS Claim 1; Page 20-21; 26pp; Japanese.
 XX
 CC This sequence is a rat CC chemokine receptor. The receptor can be used to
 CC screen for novel binding compounds and for preparation of antibodies or
 CC antiserum
 XX
 SQ Sequence 355 AA;
 XX
 Query Match 51.5%; Score 51; DB 2; Length 355;
 Best Local Similarity 66.7%; Pred. No. 3.5;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 XX
 QY 3 AHSVLSFLLWTPYAL 17
 DB 243 AITLLFFLLWTPYNL 257
 XX
 RESULT 5
 ADD45360
 ID ADD45360 standard; protein; 355 AA.
 XX
 AC ADD45360;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein NP_065417, SEQ ID NO 10793.
 XX
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.
 XX
 PN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 DR WPI; 2003-268312/26.
 DR GENBANK; NP_065417.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating


```

Query Match      49.5%; Score 49; DB 7; Length 2235;
Best Local Similarity 41.2%; Pred. No. 55;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY      3 AHSVLSFLLWTPYALKS 19
DB      1529 AYSIIAFVLRPFVNS 1545

RESULT 8
AAR67688
ID AAR67688 standard; protein; 422 AA.
XX AC AAR67688;
XX DT 16-OCT-2003 (revised)
DT 04-AUG-1995 (first entry)
XX DE S. pombe spaur1S gene product.
XX KW Aureobasidin-resistance; aureobasidin-sensitivity; spaur1S gene; mycosis;
KW antimycotic; fungicide; diagnosis; therapy.
XX OS Schizosaccharomyces pombe; JY 745.
XX PN AU9463129-A.
XX PD 01-DEC-1994.
XX PF 16-MAY-1994; 94AU-00063129.
XX PR 24-MAY-1993; 93JP-00142523.
XX PR 28-DEC-1993; 93JP-00348893.
XX PA (TAKI ) TAKARA SHUZO CO LTD.
XX PI Okada T, Takesako K, Kato I;
XX DR WPI; 1995-023286/04.
XX DR N-PSDB; AAQ75952.
XX PT New genes regulating aureobasidin sensitivity - used to develop prods.
XX PT for the diagnosis and treatment of diseases such as mycoses.
XX PS Disclosure; Page 60-62; 110pp; English.
XX CC Schizosaccharomyces pombe JY 745, which is sensitive to aureobasidin, was
CC mutagenized with EMS and a genomic library of a resistant strain was
CC prepared. A DNA fragment containing a resistant gene, spaur1R (given in
CC AAQ75952), was isolated. Using this gene as a probe, DNA containing a
CC sensitive gene, spaur1S (AAQ75953), was obtained from the sensitive
CC strain. A G1053T mutation leading to a G240C conversion gave rise to the
CC resistance. (Updated on 16-OCT-2003 to standardise OS field)
XX SQ Sequence 422 AA;

Query Match      48.5%; Score 48; DB 2; Length 422;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      4 HSVLSFLLWTPYAL 17
DB      142 HSILDILAWPYGV 155

RESULT 10
ABP55266
ID ABP55266 standard; peptide; 26 AA.
XX AC ABP55266;
XX DT 28-JAN-2003 (first entry)
XX DE Human muscarinic acetylcholine m2 receptor TM 4 peptide SEQ ID NO:19.
XX KW Human; muscarinic acetylcholine m2 receptor; histamine H2 receptor;
KW serotonin 5HT2A receptor; dopamine D2 receptor; rhodopsin; GPCR;
KW adrenergic b2 receptor; transmembrane; G protein-coupled receptor;
KW protein co-ordinate data; three dimensional structure.
XX OS Homo sapiens.
XX PN WO200279784-A1.
XX PD 10-OCT-2002.
XX PF 01-APR-2002; 2002WO-JP003264.
XX PR 30-MAR-2001; 2001JP-00101510.
XX SQ

```


CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 CC odorant, cytomagaloviral and other G-protein coupled receptors. The
 CC receptor proteins were used to design polypeptides, pref. based on the
 CC transmembrane domains, for use in G-protein coupled receptor ligand
 CC binding assays. The polypeptide fragments retain biological activity such
 CC as binding a GPR ligand or modulating GPR ligand binding to a GPR (see
 CC AAR48759-R48758, AAR50569-R50807 and AAR89189-R89195 for examples of
 CC polypeptide fragments). The polypeptide fragments can be used in
 CC compositions for treating subjects suffering from a pathology related to
 CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia
 XX
 SQ Sequence 317 AA;

Query Match 47.5%; Score 47; DB 2; Length 317;
 Best Local Similarity 58.8%; Pred. No. 15;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHVSFLSLWTPVAL 17
 DB 111 IAAWVLSFILWAPAIL 127

RESULT 13
 AAW02661
 ID AAW02661 standard; peptide; 317 AA.
 AC AAW02661;
 XX

XX 25-MAR-2003 (revised)
 DT 12-NOV-1996 (first entry)
 XX
 XX G-protein coupled human m2 muscarinic acetylcholine receptor.

XX G-protein coupled receptor; ligand binding assay; transmembrane domain;
 KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
 KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
 KW odorant; cytomagalovirus; serotonergic.
 XX

OS Homo sapiens.
 XX
 XX US5508384-A.
 PN

XX 16-APR-1996.
 PD

XX 09-SEP-1993; 93US-00118270.
 PF

XX 10-SEP-1992; 92US-00943236.
 PR

XX (UYNV) UNIV NEW YORK STATE.
 XX

XX Schuster DI, Murphy RB;
 XX

XX WPI; 1996-208785/21.
 XX

XX New dopamine receptor peptide - useful as antipsychotic agent, e.g. for
 PT treating schizophrenia.
 PT

XX Disclosure; Col 49-52; 184pp; English.
 PS

XX Proteins AAW02657-W02720 represent a range of G-protein coupled receptor
 CC (GPR) proteins selected from CAMP, adenosine, muscarinic acetylcholine,
 CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 CC odorant, cytomagaloviral and other GPR proteins. The receptor proteins
 CC were used to design polypeptides, pref. based on the transmembrane
 CC domains, for use in G-protein coupled receptor ligand binding assays. The
 CC polypeptide fragments retain biological activity such as binding a GPR
 CC ligand or modulating GPR ligand binding to a GPR (see AAW02747-W02999 for
 CC examples of polypeptide fragments). The polypeptide fragments can be used
 CC in compositions for treating subjects suffering from a pathology related
 CC to a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 XX

XX Sequence 317 AA;
 SQ

Query Match 47.5%; Score 47; DB 2; Length 317;
 Best Local Similarity 58.8%; Pred. No. 15;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHVSFLSLWTPVAL 17
 DB 111 IAAWVLSFILWAPAIL 127

RESULT 14
 ADD48620
 ID ADD48620 standard; protein; 456 AA.
 XX

XX ADD48620;
 AC

XX 29-JAN-2004 (first entry)
 DT

XX Human Protein AAK68113, SEQ ID NO 14326.
 DE

XX Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 KW

XX Homo sapiens.
 OS

XX WO2003016475-A2.
 PN

XX 27-FEB-2003.
 PD

XX 14-AUG-2002; 2002WO-US025765.
 PF

XX 14-AUG-2001; 2001US-0312147P.
 PR

XX 01-NOV-2001; 2001US-0346382P.
 PR

XX 26-NOV-2001; 2001US-0333347P.
 PR

XX (GEHO) GEN HOSPITAL CORP.
 PA

XX (FARB) BAYER AG.
 PA

XX Woolf C, D'urso D, Befort K, Costigan M;
 PI

XX WPI; 2003-268312/26.
 DR

XX GENBANK; AAK68113.
 DR

XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 PT

XX Claim 1; Page; 1017pp; English.
 PS

XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:

Search completed: July 19, 2004, 17:05:25
Job time : 61 secs

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 456 AA;

Query Match 47.5%; Score 47; DB 7; Length 456;
Best Local Similarity 58.8%; Pred. No. 22;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLIWTPTVAL 17
:|:|:|:|:|:|:|:|:|
Db 144 IAAAWVLSFILWAPAIL 160

RESULT 15
AAP96203
ID AAP96203 standard; protein; 466 AA.
XX
AC AAP96203;
XX
XX 25-MAR-2003 (revised)
DT 17-DEC-2001 (revised)
DT 27-SEP-1989 (first entry)
XX
XX Human muscarinic acetylcholine m2 receptor.
DE
XX Muscarinic acetylcholine receptor; m2; drug screening; diagnosis;
KW drug design.
XX
XX Homo sapiens.
OS
XX
XX USN7241971-N.
PN
XX
XX 14-MAR-1989.
PD
XX
XX 08-SEP-1988; 88US-00241971.
PF
XX
XX 08-SEP-1988; 88US-00241971.
PR
XX
XX (USSH) NAT INST OF HEALTH.
PA
XX
XX Bonner T;
PI
XX
XX WPI; 1989-165452/22.
DR
XX N-PSDB; AAN92069.
DR

PT Cloned genes for muscarinic acetyl:choline receptors - for drug screening
PT and diagnostic use.
XX
XX Disclosure; Page; 21pp; English.

XX Cell lines producing this MAR is useful in screening for drugs with
XX improved specificity for the part. MAR type. The amino acid sequence may
XX be useful for rational drug design and antibodies to the sequences may be
XX useful for diagnostic purposes. See also AAN92068-N92067 and AAN90086.
XX (Note: Revised entry submitted to correct the patent number format of US
XX Government-owned NTIS applications to prevent clashes with ongoing US
XX granted patent numbers. For further information please visit the Derwent
XX web site at www.derwent.com/dwpi/updates/ntis_us.html.) (Updated on 25-
XX MAR-2003 to correct PI field.)

SQ Sequence 466 AA;

Query Match 47.5%; Score 47; DB 1; Length 466;
Best Local Similarity 58.8%; Pred. No. 22;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLIWTPTVAL 17
:|:|:|:|:|:|:|:|:|
Db 144 IAAAWVLSFILWAPAIL 160


```
RESULT 5
US-08-423-752-4
; Sequence 4, Application US/08423752
; Patent No. 6022949
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; TITLE OF INVENTION: AUREOBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/423,752
; FILING DATE: April 18, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-423-752-4
Query Match 48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
Db 142 HSILDILAWPYGV 155

RESULT 6
US-08-716-873-7
; Sequence 7, Application US/08716873
; Patent No. 6194166
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-716-873-7
Query Match 48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
Db 142 HSILDILAWPYGV 155

RESULT 7
US-08-716-873-16
; Sequence 16, Application US/08716873
; Patent No. 6194166
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-716-873-16
Query Match 48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-716-873-7
Query Match 48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
Db 142 HSILDILAWPYGV 155

RESULT 7
US-08-716-873-16
; Sequence 16, Application US/08716873
; Patent No. 6194166
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-716-873-16

Query Match          48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      4 HSVLSFLWTPYAL 17
Db      142 HSILDLAWPYGV 155

RESULT 8
US-08-716-873-18
; Sequence 18, Application US/08716873
; Patent No. 6194166
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716,873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-716-873-18

Query Match          48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      4 HSVLSFLWTPYAL 17
Db      142 HSILDLAWPYGV 155

RESULT 9
US-09-368-431-7
; Sequence 7, Application US/09368431
; Patent No. 6294651
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,431
; FILING DATE: August 5, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/716,873
; FILING DATE: September 20, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-368-431-7

Query Match          48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      4 HSVLSFLWTPYAL 17
Db      142 HSILDLAWPYGV 155

RESULT 10
US-09-368-431-16
; Sequence 16, Application US/09368431
; Patent No. 6294651
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,431
; FILING DATE: August 5, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-368-431-16

Query Match          48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

; FILING DATE: August 5, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/716,873
; FILING DATE: September 20, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:

; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-368-431-16

Query Match 48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
||:| | | | |
Db 142 HSILDILAWPYGV 155

RESULT 11

US-09-368-431-18
; Sequence 18, Application US/09368431
; Patent No. 6294651
; GENERAL INFORMATION:

; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; TITLE OF INVENTION: (AS AMENDED)
; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/368,431
FILING DATE: August 5, 1999

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/716,873
FILING DATE: September 20, 1996

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 422 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

; MOLECULE TYPE: peptide
US-09-368-431-18

Query Match 48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
||:| | | | |
Db 142 HSILDILAWPYGV 155

RESULT 12

US-09-414-006-2

; Sequence 2, Application US/09414006
; Patent No. 6348577

; GENERAL INFORMATION:

; APPLICANT: Takashi OKADO et al.

; TITLE OF INVENTION: REGULATION OF AUREOBASIDIN SENSITIVITY (AS

; TITLE OF INVENTION: AMENDED)

; NUMBER OF SEQUENCES: 36

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

STREET: 2033 K Street, N.W., #800

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/414,006

FILING DATE: October 7, 1999

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/492,459

FILING DATE: June 20, 1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/243,403

FILING DATE: May 16, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200

TELEFAX: 202-721-8250
TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 422

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-414-006-2

Query Match 48.5%; Score 48; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
||:| | | | |
Db 142 HSILDILAWPYGV 155

RESULT 13

US-09-414-006-4

; Sequence 4, Application US/09414006

; Patent No. 6348577

```

;
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: REGULATION OF AUROBASIDIN SENSITIVITY (AS
; TITLE OF INVENTION: AMENDED)
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/414,006
; FILING DATE: October 7, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/492,459
; FILING DATE: June 20, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-414-006-4

Query Match 48.5%; Score 48; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSLSFLLWTPYAL 17
Db 142 HSILDILAWPYGV 155

;
; RESULT 14
; US-09-447-223-2
; Sequence 2, Application US/09447223
; Patent No. 6432664
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,223
; FILING DATE: 23-No. 6432664-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: <Unknown>
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS

```

```

;
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,223
; FILING DATE: 23-No. 6432664-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: <Unknown>
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
; US-09-447-223-2

Query Match 48.5%; Score 48; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSLSFLLWTPYAL 17
Db 142 HSILDILAWPYGV 155

;
; RESULT 15
; US-09-447-223-4
; Sequence 4, Application US/09447223
; Patent No. 6432664
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,223
; FILING DATE: 23-No. 6432664-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: <Unknown>
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

```

;
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-447-223-4

Query Match 48.5%; Score 48; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSLFSLMTPTAL 17
||:| | | | :
Db 142 HSILDLAWPYGV 155

Search completed: July 19, 2004, 17:07:25
Job time : 20 secs

This Page Blank (Rev. 11/11)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:06:36 ; Search time 43 Seconds

(without alignments)
145.377 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99

Sequence: 1 MAHSVLSFLLWTPYALKSX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pbp:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pbp:*
- 3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pbp:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pbp:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pbp:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pbp:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pbp:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pbp:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pbp:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pbp:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pbp:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pbp:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pbp:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pbp:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pbp:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pbp:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pbp:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pbp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	99.0	20	14	US-10-195-730-116
2	51	51.5	355	14	US-10-376-564-13
3	49	49.5	364	16	US-10-437-963-163690
4	48	48.5	422	10	US-09-951-217-7
5	48	48.5	422	10	US-09-951-217-16
6	48	48.5	422	10	US-09-951-217-18
7	47	47.5	196	14	US-10-017-161-794
8	47	47.5	466	11	US-09-826-509-515
9	47	47.5	466	14	US-10-225-567A-190
10	47	47.5	488	13	US-10-029-009-7
11	47	47.5	496	13	US-10-029-009-34
12	47	47.5	500	13	US-10-029-009-19
13	46.5	47.0	32	15	US-10-242-355-332
14	46	46.5	345	14	US-10-017-161-542
15	46	46.5	345	15	US-10-292-798-476

16	46	46.5	478	14	US-10-225-567A-193	Sequence 193, App
17	46	46.5	479	11	US-09-826-509-519	Sequence 519, App
18	46	46.5	490	10	US-09-992-238-7	Sequence 7, Appli
19	46	46.5	1907	14	US-10-156-761-10430	Sequence 10430, A
20	45.5	46.0	681	15	US-10-369-493-13394	Sequence 13394, A
21	45	45.5	154	16	US-10-437-963-119491	Sequence 119491,
22	45	45.5	209	9	US-09-925-302-752	Sequence 752, App
23	45	45.5	209	12	US-09-925-302-752	Sequence 752, App
24	45	45.5	279	12	US-10-246-583-58	Sequence 58, Appl
25	45	45.5	279	15	US-10-174-364-58	Sequence 58, Appl
26	45	45.5	355	9	US-09-961-068-1	Sequence 1, Appli
27	45	45.5	355	9	US-09-960-547-1	Sequence 1, Appli
28	45	45.5	355	12	US-09-893-512-11	Sequence 11, Appl
29	45	45.5	355	12	US-10-039-659-13	Sequence 13, Appl
30	45	45.5	355	14	US-10-225-567A-62	Sequence 62, Appl
31	45	45.5	355	14	US-10-245-850-2	Sequence 2, Appli
32	45	45.5	355	14	US-10-164-649-53	Sequence 53, Appl
33	45	45.5	355	14	US-10-376-564-14	Sequence 14, Appl
34	45	45.5	355	14	US-10-239-423-62	Sequence 62, Appl
35	45	45.5	355	14	US-10-439-845-9	Sequence 9, Appli
36	45	45.5	355	15	US-10-452-015-2	Sequence 2, Appli
37	45	45.5	375	14	US-10-219-834-78	Sequence 78, Appl
38	45	45.5	500	16	US-10-437-963-154068	Sequence 154068,
39	44	44.4	771	12	US-10-425-114-60213	Sequence 60213, A
40	44	44.4	824	12	US-10-282-122A-52599	Sequence 52599, A
41	43	43.4	68	12	US-10-424-599-176608	Sequence 176608,
42	43	43.4	102	12	US-10-424-599-172079	Sequence 172079,
43	43	43.4	141	12	US-10-424-599-188984	Sequence 188984,
44	43	43.4	246	15	US-10-369-493-12318	Sequence 12318, A
45	43	43.4	284	9	US-09-867-550-1872	Sequence 1872, Ap

ALIGNMENTS

RESULT 1

US-10-195-730-116
; Sequence 116, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 116
; LENGTH: 20
; TYPE: ERT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (20)
; OTHER INFORMATION: Xaa equals stop translation
US-10-195-730-116

Query Match 99.0%; Score 98; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.7e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLLWTPYALKS 19
Db 1 MAHSVLSFLLWTPYALKS 19

RESULT 2

US-10-376-564-13
; Sequence 13, Application US/10376564
; Publication No. US20030180302A1
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckhard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014003
; CURRENT APPLICATION NUMBER: US/10/376,564
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 09/886,319
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 100 30 149.5-41
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-376-564-13
Query Match 51.5%; Score 51; DB 14; Length 355;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 3 AHSVLSFLWTPYAL 17
Db 243 AITLLFLLWTPYNL 257
RESULT 3
US-10-437-963-163690
; Sequence 163690, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 163690
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(364)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_62662C.1.pep
US-10-437-963-163690
Query Match 49.5%; Score 49; DB 16; Length 364;
Best Local Similarity 62.5%; Pred. No. 27;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYA 16
Db 148 MAANDGLSLLWMPYS 163
RESULT 4
US-09-951-217-7
; Sequence 7, Application US/09951217
; Publication No. US20030050440A1
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; (AS AMENDED)
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WENDROTH, LIND & PONACK, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/951,217
; FILING DATE: 06-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/368,431
; FILING DATE: August 5, 1999
; APPLICATION NUMBER: 08/716,873
; FILING DATE: September 20, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-951-217-7
Query Match 48.5%; Score 48; DB 10; Length 422;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy 4 HSVLSFLWTPYAL 17
Db 142 HSLDILAWPYGV 155
RESULT 5
US-09-951-217-16
; Sequence 16, Application US/09951217
; Publication No. US20030050440A1
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; (AS AMENDED)
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:

ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/951,217
FILING DATE: 06-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/368,431
FILING DATE: August 5, 1999
APPLICATION NUMBER: 08/716,873
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-951-217-16
Query Match 48.5%; Score 48; DB 10; Length 422;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 4 HSVLSFLWTPYAL 17
DB 142 HSILDILAWPYGV 155
RESULT 6
US-09-951-217-18
Sequence 18, Application US/09951217
Publication No. US2003005040A1
GENERAL INFORMATION:
APPLICANT: Takashi OKADO et al.
TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
(AS AMENDED)
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/951,217
FILING DATE: 06-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/368,431
FILING DATE: August 5, 1999
APPLICATION NUMBER: 08/716,873
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-951-217-18
Query Match 48.5%; Score 48; DB 10; Length 422;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 4 HSVLSFLWTPYAL 17
DB 142 HSILDILAWPYGV 155
RESULT 7
US-10-017-161-794
Sequence 794, Application US/10017161
Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUWA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017,161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 794
LENGTH: 196
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: MOD_RES
LOCATION: (25)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (52)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (59)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (67)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (73)
OTHER INFORMATION: Variable amino acid
FEATURE:

NAME/KEY: MOD RES
LOCATION: (78)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (101)..(104)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (126)..(127)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (134)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (147)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (166)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (172)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (179)
OTHER INFORMATION: Variable amino acid
US-10-017-161-794

Query Match 47.5%; Score 47; DB 14; Length 196;
Best Local Similarity 58.8%; Pred. No. 30;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLLWTPYAL 17
: || ||||| :
Db 3 IAAWVLSFILWAPAIL 19

RESULT 8
US-09-826-509-515
Sequence 515, Application US/09826509
Publication No. US20030204073A1
GENERAL INFORMATION:
APPLICANT: Lehmann-Bruinsma, Karin
APPLICANT: Liaw, Chen W.
TITLE OF INVENTION: No. US20030204073A1-Endogenous, Constitutively Activated Known G
TITLE OF INVENTION: Protein-Coupled Receptors
FILE REFERENCE: AREN-207
CURRENT APPLICATION NUMBER: US/09/826,509
CURRENT FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/195,747
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: 09/170,496
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 589
SOFTWARE: PatentIn Version 2.1
SEQ ID NO 515
LENGTH: 466
TYPE: PRT
ORGANISM: Homo sapiens
US-09-826-509-515

Query Match 47.5%; Score 47; DB 11; Length 466;
Best Local Similarity 58.8%; Pred. No. 68;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLLWTPYAL 17
: || ||||| :
Db 3 IAAWVLSFILWAPAIL 19

Db 144 IAAWVLSFILWAPAIL 160

RESULT 9

US-10-225-567A-190
Sequence 190, Application US/10225567A
Publication No. US20030113798A1
GENERAL INFORMATION:
APPLICANT: LifeSpan Biosciences
APPLICANT: Brown, Joseph P.
APPLICANT: Burmer, Glenn C.
APPLICANT: Roush, Christine L.
TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTOR
FILE REFERENCE: 1920-4-4
CURRENT APPLICATION NUMBER: US/10/225,567A
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/257,144
PRIOR FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 2292
SOFTWARE: PatentIn version 3.1
SEQ ID NO 190
LENGTH: 466
TYPE: PRT
ORGANISM: Homo sapiens
US-10-225-567A-190

Query Match 47.5%; Score 47; DB 14; Length 466;
Best Local Similarity 58.8%; Pred. No. 68;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLLWTPYAL 17
: || ||||| :
Db 144 IAAWVLSFILWAPAIL 160

RESULT 10

US-10-029-009-7
Sequence 7, Application US/10029009
Publication No. US20020164617A1
GENERAL INFORMATION:
APPLICANT: Felsch, Jason S.
APPLICANT: Annis, David Allen
APPLICANT: Kalghatgi, Krishna
APPLICANT: Nash, Huw M.
TITLE OF INVENTION: Affinity Selection-Based Screening of Hydrophobic Proteins
FILE REFERENCE: 111025.173 US2
CURRENT APPLICATION NUMBER: US/10/029,009
CURRENT FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: US 60/258,970
PRIOR FILING DATE: 2000-12-29
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 488
TYPE: PRT
ORGANISM: Human
US-10-029-009-7

Query Match 47.5%; Score 47; DB 13; Length 488;
Best Local Similarity 58.8%; Pred. No. 71;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLLWTPYAL 17
: || ||||| :
Db 166 IAAWVLSFILWAPAIL 182

RESULT 11

US-10-029-009-34
Sequence 34, Application US/10029009
Publication No. US20020164617A1
GENERAL INFORMATION:
APPLICANT: Felsch, Jason S.

```

; APPLICANT: Annis, David Allen
; APPLICANT: Kalghatgi, Krishna
; APPLICANT: Nash, Huw M.
; TITLE OF INVENTION: Affinity Selection-Based Screening of Hydrophobic Proteins
; FILE REFERENCE: 111025.173 US2
; CURRENT APPLICATION NUMBER: US/10/029,009
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/258,970
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Met-Melitin SS-FLAG M1 tag-m2 mChR sequence
US-10-029-009-34

Query Match          47.5%; Score 47; DB 13; Length 496;
Best Local Similarity 58.8%; Pred. No. 72;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 MAHSVLSFLWTPYAL 17
Db      174 IAAAWLSFILWAPAIL 190

RESULT 12
US-10-029-009-19
; Sequence 19, Application US/10029009
; Publication No. US20020164617A1
; GENERAL INFORMATION:
; APPLICANT: Felsch, Jason S.
; APPLICANT: Annis, David Allen
; APPLICANT: Kalghatgi, Krishna
; APPLICANT: Nash, Huw M.
; TITLE OF INVENTION: Affinity Selection-Based Screening of Hydrophobic Proteins
; CURRENT APPLICATION NUMBER: US/10/029,009
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/258,970
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Human
US-10-029-009-19

Query Match          47.5%; Score 47; DB 13; Length 500;
Best Local Similarity 58.8%; Pred. No. 73;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 MAHSVLSFLWTPYAL 17
Db      184 IAAAWLSFILWAPAIL 200

RESULT 13
US-10-242-355-332
; Sequence 332, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065

```

```

; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 332
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (26)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-242-355-332

Query Match          47.0%; Score 46.5; DB 15; Length 32;
Best Local Similarity 56.2%; Pred. No. 6.3;
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY      1 MAHSVLSFLWTPY 15
Db      17 MAPHTLMSFLCWLPY 32

RESULT 14
US-10-017-161-542
; Sequence 542, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 542
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (4)..(37)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (49)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (271)
; OTHER INFORMATION: Variable amino acid
; FEATURE:

```

; NAME/KEY: MOD RES
; LOCATION: (276)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (285)..(318)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (330)
; OTHER INFORMATION: Variable amino acid
US-10-017-161-542

Query Match 46.5%; Score 46; DB 14; Length 345;

Best Local Similarity 58.8%; Pred. No. 72; Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLWTPYAL 17
Db 150 IAAAVLSFVWAPAIL 166

RESULT 15

US-10-292-798-476
; Sequence 476, Application US/10292798
; Publication No. US2003023583A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 476
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (4)..(37)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (49)..(49)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (271)..(271)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (276)..(276)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (285)..(318)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (330)..(330)
; OTHER INFORMATION: Variable amino acid
US-10-292-798-476

Query Match 46.5%; Score 46; DB 15; Length 345;

Best Local Similarity 58.8%; Pred. No. 72;

Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAAHSVLSFLWTPYAL 17
Db 150 IAAAVLSFVWAPAIL 166
Search completed: July 19, 2004, 17:12:04
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:03:23 ; Search time 16 Seconds
(without alignments)
120.239 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78:*

2: PIR1:*

3: PIR2:*

4: PIR3:*

5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	51.5	355	2 I49339	macrophage inflamm
2	49	49.5	2124	2 T01526	hypothetical prote
3	48	48.5	422	2 T38764	aureobasidin a res
4	47.5	48.0	355	2 I51319	RH2 opsin - green
5	47	47.5	466	2 A40972	muscarinic acetyl
6	47	47.5	466	2 JH0197	muscarinic acetyl
7	47	47.5	466	2 S10126	muscarinic acetyl
8	47	47.5	466	2 S10836	muscarinic acetyl
9	47	47.5	466	2 A27386	muscarinic acetyl
10	46	46.5	478	2 C29514	muscarinic acetyl
11	46	46.5	479	2 S10127	muscarinic acetyl
12	46	46.5	479	2 S33776	muscarinic acetyl
13	46	46.5	490	2 A35546	muscarinic acetyl
14	45.5	46.0	355	1 A46191	iodopsin homolog -
15	45	45.5	355	2 A45177	chemokine (C-C) re
16	44.5	44.9	355	2 A42347	opsin, green-sensi
17	44	44.4	301	2 T33218	hypothetical prote
18	44	44.4	308	2 T36751	hypothetical prote
19	44	44.4	557	2 S21733	PACC protein - hum
20	44	44.4	563	2 T14242	NADH2 dehydrogenas
21	44	44.4	649	2 B38129	bo-type ubiquinol
22	43	43.4	354	2 I58186	probable G protein
23	43	43.4	414	2 AF3423	cyanate transport
24	43	43.4	459	2 T11149	NADH2 dehydrogenas
25	43	43.4	870	2 T30110	hypothetical prote
26	42.5	42.9	349	2 B45229	opsin, green-sensi
27	42	42.4	281	2 S54137	probable movement
28	42	42.4	305	2 T45942	hypothetical prote
29	42	42.4	305	2 A84799	hypothetical prote

30 42 42.4 355 2 JC4304 orphan G protein-c
31 42 42.4 359 2 I49341 MIP-1 alpha recept
32 42 42.4 360 2 JC2443 chemokine (C-C) re
33 42 42.4 374 2 I38450 chemokine (C-C) re
34 42 42.4 589 2 AH0370 probable binding p
35 42 42.4 729 2 A91038 hypothetical prote
36 42 42.4 742 2 D85882 hypothetical prote
37 42 42.4 768 2 H65013 yfeA protein - Esc
38 42 42.4 772 2 H85054 hypothetical prote
39 41 41.4 278 2 B1601 ABC transporter, p
40 41 41.4 278 2 F72064 solute binding pro
41 41 41.4 278 2 A86558 solute binding pro
42 41 41.4 310 2 G86490 probable bionin ho
43 41 41.4 323 2 B83215 conserved hypothet
44 41 41.4 376 2 S28853 opsin - red swamp
45 41 41.4 378 2 S28789 opsin Rh4 - fruit

ALIGNMENTS

RESULT 1

I49339
macrophage inflammatory protein-1 alpha receptor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49339
R:Gao, J.L.; Murphy, P.M.
J. Biol. Chem. 270, 17494-17501, 1995
A:Title: Cloning and differential tissue-specific expression of three mouse beta chemoki-
A:Reference number: I49339; MUID:95340546; PMID:7542241
A:Accession: I49339
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-355 <RES>
A:Cross-references: EMBL:U28404; NID:g881547; PIDN:AAA89153.1; PID:g881548
C:Superfamily: vertebrate rhodopsin

Query Match 51.5%; Score 51; DB 2; Length 355;
Best Local Similarity 66.7%; Pred. No. 1.3;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AHSVLSFLWTPYAL 17
Db 243 AITLFFLLWTPYNL 257

RESULT 2

T01526
hypothetical protein A_IG005I10.24 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Nov-1999
C:Accession: T01526
R:Andrews, S.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of A. thaliana IG005I10.
A:Reference number: Z14347
A:Accession: T01526
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2124 <AND>
A:Cross-references: EMBL:AF013293; NID:g2252823; PID:g2252843
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 11/3; 49/3; 84/3; 141/2; 256/2; 989/3; 1019/2; 1889/2; 1951/2; 1973/3
A:Note: A_IG005I10.24
C:Superfamily: Arabidopsis thaliana hypothetical protein A_IG005I10.24

Query Match 49.5%; Score 49; DB 2; Length 2124;
Best Local Similarity 41.2%; Pred. No. 16;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHSVLSFLLWTPYALKS 19
 ||:|:|:|:|:|:|
 Db 1438 AYSIIAFVLWRPFVNS 1454

RESULT 3
 T38764
 aureobasidin a resistance protein - fission yeast (Schizosaccharomyces pombe)
 C;Species: Schizosaccharomyces pombe
 C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C;Accession: T38764
 R;Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, January 1996
 A;Reference number: Z21810
 A;Accession: T38764
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-422 <GEN>
 A;Cross-references: EMBL:Z69086; PIDN:CAA93163.1; GSPDB:GN000066; SPDB:SPAC3H8.06
 A;Experimental source: strain 972h-; cosmid c3H8
 C;Genetics:
 A;Gene: aur1; SPDB:SPAC3H8.06
 A;Map position: 1

Query Match 48.5%; Score 48; DB 2; Length 422;
 Best Local Similarity 50.0%; Pred. No. 4.7;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLLWTPYAL 17
 ||:|:|:|:|:|:|
 Db 142 HSIIDILAWPYGV 155

RESULT 4
 I51319
 RH2 opsin - green anole
 C;Species: Anolis carolinensis (green anole)
 C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
 C;Accession: I51319
 R;Kawamura, S.; Yokoyama, S.
 J. Mol. Evol. 40, 594-600, 1995
 A;Title: Paralogous origin of the rhodopsinlike opsin genes in lizards.
 A;Reference number: I51319; MUID:95371134; PMID:7643409
 A;Accession: I51319
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-355 <KAW>
 A;Cross-references: GB:S79167; NID:gl042072; PIDN:AAB35062.1; PID:gl042073
 C;Genetics:
 A;Gene: rh2Ac
 A;Introns: 121/1; 177/2; 232/3; 312/3
 C;Superfamily: vertebrate rhodopsin
 C;Keywords: chromoprotein; eye; G protein-coupled receptor; lipoprotein; phosphoprotein;
 F;296/Binding site: retinal (lys) (covalent) #status predicted
 F;322,323/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 48.0%; Score 47.5; DB 2; Length 355;
 Best Local Similarity 76.3%; Pred. No. 4.8;
 Matches 10; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 6 VLSFLL-WTPYAL 17
 ||:|:|:|:|:|:|
 Db 258 VLGFLLAWTPYAM 270

RESULT 5
 A40972
 muscarinic acetylcholine receptor M2 - chicken
 C;Species: Gallus gallus (chicken)
 C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 24-Nov-1999
 C;Accession: A40972
 R;Rietje, K.M.; Nathanson, N.M.
 J. Biol. Chem. 266, 17382-17387, 1991

A;Title: Embryonic chick heart expresses multiple muscarinic acetylcholine receptor subunit
 affinity for pirenzepine.
 A;Reference number: A40972; MUID:91373358; PMID:1840593
 A;Accession: A40972
 A;Molecule type: DNA
 A;Residues: 1-466 <TIE>
 A;Cross-references: GB:M73217
 C;Genetics:
 A;Gene: CM2
 A;Introns: #status absent
 C;Superfamily: vertebrate rhodopsin
 C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phospho
 F;26-51/Domain: transmembrane #status predicted <TM1>
 F;64-88/Domain: transmembrane #status predicted <TM2>
 F;101-122/Domain: transmembrane #status predicted <TM3>
 F;142-165/Domain: transmembrane #status predicted <TM4>
 F;187-210/Domain: transmembrane #status predicted <TM5>
 F;389-409/Domain: transmembrane #status predicted <TM6>
 F;421-442/Domain: transmembrane #status predicted <TM7>

Query Match 47.5%; Score 47; DB 2; Length 466;
 Best Local Similarity 58.8%; Pred. No. 7.6;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLLWTPYAL 17
 ||:|:|:|:|:|:|
 Db 147 IAAAVLSFILWAPAIL 163

RESULT 6
 JH0197
 muscarinic acetylcholine receptor M2 - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 24-Nov-1999
 C;Accession: JH0197; D37121
 R;Jai, J.; Bloom, J.W.; Yamamura, H.I.; Roeske, W.R.
 Life Sci. 47, 1001-1013, 1990
 A;Title: Amplification of the rat m2 muscarinic receptor gene by the polymerase chain re
 A;Reference number: JH0197; MUID:91041524; PMID:2172674
 A;Accession: JH0197
 A;Molecule type: DNA
 A;Residues: 1-466 <JAI>
 R;Kurtenbach, E.; Curtis, C.A.M.; Pedder, E.K.; Aitken, A.; Harris, A.C.M.; Hulme, E.C.
 J. Biol. Chem. 265, 13702-13708, 1990
 A;Title: Muscarinic acetylcholine receptors. Peptide sequencing identifies residues invo
 A;Reference number: A37121; MUID:90337982; PMID:2380182
 A;Accession: D37121
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 60-122 <KUR>
 C;Superfamily: vertebrate rhodopsin
 C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; transme
 F;23-48/Domain: transmembrane #status predicted <TM1>
 F;61-85/Domain: transmembrane #status predicted <TM2>
 F;98-119/Domain: transmembrane #status predicted <TM3>
 F;139-162/Domain: transmembrane #status predicted <TM4>
 F;184-207/Domain: transmembrane #status predicted <TM5>
 F;389-409/Domain: transmembrane #status predicted <TM6>
 F;421-442/Domain: transmembrane #status predicted <TM7>

Query Match 47.5%; Score 47; DB 2; Length 466;
 Best Local Similarity 58.8%; Pred. No. 7.6;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLLWTPYAL 17
 ||:|:|:|:|:|:|
 Db 144 IAAAVLSFILWAPAIL 160

RESULT 7
 S10126
 muscarinic acetylcholine receptor M2 - human
 C;Species: Homo sapiens (man)

C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 24-Nov-1999
A;Accession: S10126
R;Peralta, E.G.; Ashkenazi, A.; Winslow, J.W.; Smith, D.H.; Ramachandran, J.; Capon, D.J.
EMBO J. 6, 3923-3929, 1987
A;Title: Distinct primary structures, ligand-binding properties and tissue-specific expression of the human interleukin-3 receptor
A;Reference number: S04326; MUID:88166632; PMID:3443095
A;Accession: S10126
A;Molecule type: DNA
A;Residues: 1-466 <PER>
A;Cross-references: EMBL:X15264; NID:g32319; PIDN:CAA33335.1; PID:g32320
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phosphatase
F;23-48/Domain: transmembrane #status predicted <TM1>
F;60-89/Domain: transmembrane #status predicted <TM2>
F;98-119/Domain: transmembrane #status predicted <TM3>
F;139-162/Domain: transmembrane #status predicted <TM4>
F;184-207/Domain: transmembrane #status predicted <TM5>
F;389-409/Domain: transmembrane #status predicted <TM6>
F;421-442/Domain: transmembrane #status predicted <TM7>
F;2,6,9/Binding site: carboxylate (Asn) (covalent) #status predicted
Query Match 47.5%; Score 47; DB 2; Length 466;
Best Local Similarity 58.8%; Pred. No. 7.6;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSVLSFLWTPYAL 17
Db 144 TAAAWLSFILWAPAIL 160
RESULT 8
S10856
muscarinic acetylcholine receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 24-Nov-1999
C;Accession: S10856
R;Gocayne, J.; Robinson, D.A.; FitzGerald, M.G.; Chung, F.Z.; Kerlavage, A.R.; Lentes, K.
Proc. Natl. Acad. Sci. U.S.A. 84, 8296-8300, 1987
A;Title: Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic receptors
A;Reference number: S10855; MUID:88068581; PMID:2825184
A;Accession: S10856
A;Molecule type: mRNA
A;Residues: 1-466 <GOC>
A;Cross-references: EMBL:J03025; NID:g203461; PIDN:AAA40926.1; PID:g203462
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phosphatase
F;23-48/Domain: transmembrane #status predicted <TM1>
F;61-85/Domain: transmembrane #status predicted <TM2>
F;98-119/Domain: transmembrane #status predicted <TM3>
F;139-162/Domain: transmembrane #status predicted <TM4>
F;184-207/Domain: transmembrane #status predicted <TM5>
F;389-409/Domain: transmembrane #status predicted <TM6>
F;421-442/Domain: transmembrane #status predicted <TM7>
Query Match 47.5%; Score 47; DB 2; Length 466;
Best Local Similarity 58.8%; Pred. No. 7.6;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSVLSFLWTPYAL 17
Db 144 TAAAWLSFILWAPAIL 160
RESULT 9
A27386
muscarinic acetylcholine receptor, cardiac - pig
N;Alternate names: muscarinic acetylcholine receptor M2
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 31-Dec-2000
A;Accession: A27386; A25656
R;Peralta, E.G.; Winslow, J.W.; Peterson, G.L.; Smith, D.H.; Ashkenazi, A.; Ramachandran, J.
Science 236, 600-605, 1987
A;Title: Primary structure and biochemical properties of an M-2 muscarinic receptor.

A;Reference number: A27386; MUID:87206169; PMID:3107123
A;Accession: A27386
A;Molecule type: DNA; mRNA
A;Residues: 1-466 <PER>
A;Cross-references: GB:M16331; NID:g164311; PIDN:AAA30986.1; PID:g164313
A;Experimental source: atrial muscle
A;Note: the protein sequence derived from the mRNA clones differs from that of the genomic sequence
R;Castro, J.
FEBS Lett. 209, 367-372, 1986
A;Title: Primary structure of porcine cardiac muscarinic acetylcholine receptor deduced from complementary DNA
A;Reference number: A25656; MUID:87080790; PMID:3792556
A;Accession: A25656
A;Molecule type: mRNA
A;Residues: 1-329, 'K', 331-466 <KUB>
A;Cross-references: GB:X04708; NID:g1859; PIDN:CAA28413.1; PID:g1860
A;Experimental source: cardiac muscle
C;Superfamily: vertebrate rhodopsin
C;Keywords: cardiac muscle; G protein-coupled receptor; glycoprotein; heart; neurotransmitter
F;23-48/Domain: transmembrane #status predicted <TM1>
F;61-85/Domain: transmembrane #status predicted <TM2>
F;98-119/Domain: transmembrane #status predicted <TM3>
F;139-162/Domain: transmembrane #status predicted <TM4>
F;184-207/Domain: transmembrane #status predicted <TM5>
F;389-409/Domain: transmembrane #status predicted <TM6>
F;421-442/Domain: transmembrane #status predicted <TM7>
Query Match 47.5%; Score 47; DB 2; Length 466;
Best Local Similarity 58.8%; Pred. No. 7.6;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSVLSFLWTPYAL 17
Db 144 TAAAWLSFILWAPAIL 160
RESULT 10
C29514
muscarinic acetylcholine receptor M4 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 24-Nov-1999
C;Accession: C94518; C94293; E37121; C29514
R;Bonner, T.I.
submitted to GenBank, July 1987
A;Reference number: A94518
A;Accession: C94518
A;Molecule type: mRNA
A;Residues: 1-478 <BO1>
R;Bonner, T.I.; Buckley, N.J.; Young, A.C.; Brann, M.R.
Science 237, 527-532, 1987
A;Title: Identification of a family of muscarinic acetylcholine receptor genes.
A;Reference number: A94293; MUID:87263421; PMID:3037705
A;Accession: C94293
A;Molecule type: mRNA
A;Residues: 1-233;373-478 <BO2>
A;Experimental source: cerebral cortex
A;Note: only a part of the protein translation is given; none of the nucleotide sequence
R;Kuttenbach, E.; Curtis, C.A.M.; Pedder, E.K.; Aiken, A.; Harris, A.C.M.; Hulme, E.C.
J. Biol. Chem. 265, 13702-13708, 1990
A;Title: Muscarinic acetylcholine receptors. Peptide sequencing identifies residues involved in binding of the antagonist atropine
A;Reference number: A37121; MUID:90337982; PMID:2380182
A;Accession: E37121
A;Status: preliminary
A;Molecule type: protein
A;Residues: 68-130 <KUR>
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phosphatase
F;31-56/Domain: transmembrane #status predicted <TM1>
F;69-93/Domain: transmembrane #status predicted <TM2>
F;106-127/Domain: transmembrane #status predicted <TM3>
F;147-170/Domain: transmembrane #status predicted <TM4>
F;192-215/Domain: transmembrane #status predicted <TM5>
F;401-421/Domain: transmembrane #status predicted <TM6>
F;433-454/Domain: transmembrane #status predicted <TM7>

ATTENTION: C-2 CNA-1, INACT

chemokine (C-C) receptor 1 - human
N; Alternate names: C-C CKR-1; macrophage inflammatory protein-1-alpha receptor

C:Species: Homo sapiens (man)
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999
 C:Accession: A45177; I55671
 R:Neote, K.; DiGregorio, D.; Mak, J.Y.; Horuk, R.; Schall, T.J.
 Cell 72, 415-425, 1993
 A:Title: Molecular cloning, functional expression, and signaling characteristics of a C-
 terminal domain of the human macrophage inflammatory protein 1 alpha
 A:Reference number: A45177; MUID:93161416; PMID:7679328
 A:Accession: A45177
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-355 <NEO>
 A:Cross-references: GB:L10918; NID:g292416; PIDN:AAA36543.1; PID:g292417
 A:Experimental source: HL60 cells
 A>Note: sequence extracted from NCBI backbone (NCBIP:124876)
 R:Gao, J.
 J. Exp. Med. 177, 1421-1427, 1993
 A:Title: Structure and functional expression of the human macrophage inflammatory 1 alpha
 A:Reference number: I55671; MUID:93240122; PMID:7683036
 A:Accession: I55671
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-355 <RES>
 A:Cross-references: GB:L10918; NID:g292416; PIDN:AAA36543.1; PID:g292417
 C:Genetics:
 A:Gene: GDB:CMKBR1; CMKR-1
 A:Cross-references: GDB:138446; OMIM:601159
 A:Map position: 3p21-3p21
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: disulfide bond; G protein-coupled receptor; glycoprotein; phosphoprotein; tr
 F:36-60/Domain: transmembrane #status predicted <TM1>
 F:71-91/Domain: transmembrane #status predicted <TM2>
 F:108-129/Domain: transmembrane #status predicted <TM3>
 F:147-171/Domain: transmembrane #status predicted <TM4>
 F:205-223/Domain: transmembrane #status predicted <TM5>
 F:240-264/Domain: transmembrane #status predicted <TM6>
 F:288-305/Domain: transmembrane #status predicted <TM7>
 F:5/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:24-273,106-183/disulfide bonds: #status predicted
 F:345/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted

Query Match 45.5%; Score 45; DB 2; Length 355;
 Best Local Similarity 58.3%; Pred. No. 12;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 6 VLSFLWTPYAL 17
 Db 246 IIFFLWTPYNL 257

Search completed: July 19, 2004, 17:06:56
 Job time : 17 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 16:59:32 ; Search time 13 Seconds
(without alignments)
80.108 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99
Sequence: 1 MAHVSFLWTFYALKSX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	51.5	355	1	CKR1_MOUSE
2	48	48.5	422	1	AUR1_SCHPO
3	47.5	48.0	355	1	OPSB_ANOCA
4	47	47.5	377	1	OPSD_HEMSA
5	47	47.5	440	1	ACM2_PANTR
6	47	47.5	466	1	ACM2_CHICK
7	47	47.5	466	1	ACM2_HUMAN
8	47	47.5	466	1	ACM2_MOUSE
9	47	47.5	466	1	ACM2_PIG
10	47	47.5	466	1	ACM2_RAT
11	46	46.5	478	1	ACM4_RAT
12	46	46.5	479	1	ACM4_HUMAN
13	46	46.5	479	1	ACM4_MOUSE
14	46	46.5	490	1	ACM4_CHICK
15	45.5	46.0	355	1	OPSB_GEGE
16	45	45.5	355	1	CKR1_HUMAN
17	45	45.5	355	1	CKR1_MACMU
18	45	45.5	382	1	CKD6_RAT
19	44.5	44.9	351	1	OPSG_CHICK
20	44	44.4	558	1	FACC_HUMAN
21	44	44.4	649	1	COX1_BACSU
22	43.5	43.9	349	1	OPSH_BRARE
23	43.5	43.9	399	1	OPSD_PATYE
24	43	43.4	354	1	C3X1_RAT
25	43	43.4	498	1	Y507_CAEEL
26	42.5	42.9	349	1	OPSH_CARAU
27	42	42.4	176	1	IL20_MOUSE
28	42	42.4	354	1	CKR5_MOUSE
29	42	42.4	354	1	CKR5_RAT
30	42	42.4	355	1	C3X1_HUMAN
31	42	42.4	358	1	CKR3_CAVPO
32	42	42.4	359	1	CKR3_MOUSE
33	42	42.4	359	1	CKR3_RAT

34	42	42.4	360	1	CKR2_MACMU
35	42	42.4	373	1	CKR2_MOUSE
36	42	42.4	373	1	CKR2_RAT
37	42	42.4	374	1	CKR2_HUMAN
38	42	42.4	377	1	OPSI_HEMSA
39	42	42.4	378	1	CKD6_MOUSE
40	42	42.4	729	1	YFEA_ECOLI
41	41.5	41.9	289	1	OPSD_COTIN
42	41	41.4	278	1	Y541_CHLPN
43	41	41.4	298	1	OPSD_PROOR
44	41	41.4	300	1	OPSD_CAMMA
45	41	41.4	301	1	OPSD_CAMLU

ALIGNMENTS

RESULT 1
CKR1_MOUSE
ID CKR1_MOUSE STANDARD; PRT; 355 AA.
AC PS1675; Q91VP9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE C-C chemokine receptor type 1 (C-C CKR-1) (CC-CKR-1) (CCR-1) (CCR1)
DE (Macrophage inflammatory protein-1 alpha receptor) (MIP-lalpha-R)
DE (RANTES-R).
GN CCR1 OR CMKCR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Peritoneal macrophage;
RX MEDLINE=96072806; PubMed=7594543;
RA Post T.W., Bozic C.R., Rothenberg M.E., Luster A.D., Gerard N.,
RA Gerard C.;
RT "Molecular characterization of two murine eosinophil beta chemokine
receptors.";
RT J. Immunol. 155:5299-5305 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=95340546; PubMed=7542241;
RA Gao J.-L., Murphy P.M.;
RT "Cloning and differential tissue-specific expression of three mouse
beta chemokine receptor-like genes, including the gene for a
functional macrophage inflammatory protein-1 alpha receptor.";
RT J. Biol. Chem. 270:17494-17501 (1995).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McWeeney P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fabey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";

```

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
-!- FUNCTION: RECEPTOR FOR A C-C TYPE CHEMOKINE. BINDS TO MIP-1-ALPHA,
RANTES, AND LESS EFFICIENTLY, TO MIP-1-BETA OR MCP-1 AND
SUBSEQUENTLY TRANSDUCES A SIGNAL BY INCREASING THE INTRACELLULAR
CALCIUM IONS LEVEL. RESPONSIBLE FOR AFFECTING STEM CELL
PROLIFERATION.
-!- SUBCELLULAR LOCATION: Integral membrane protein.
-!- TISSUE SPECIFICITY: Detected in the heart, spleen, lung,
peritoneal exudate cells and leukocytes.
-!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U29678; AAA86119.1; -.
EMBL; U28404; AAA89153.1; -.
EMBL; BC011092; AAH11092.1; -.
PIR; I49339; I49339; Ccrl.
MGD; MGI:104618; Ccrl.
GO; GO:0016493; F:C-C chemokine receptor activity; IDA.
GO; GO:0005155; F:protein binding; IPI.
GO; GO:00030595; P:immune cell chemotaxis; IDA.
GO; GO:0006954; P:inflammatory response; IMP.
GO; GO:0030099; P:myeloid blood cell differentiation; IMP.
InterPro; IPR000276; GPCR_Rhodopsin.
PFam; PF00001; 7tm 1; 1.
PRINTS; PR00237; GPCRHOPOPS.
PROSITE; PS00237; G PROTEIN RECEPTOR FL 1; 1.
PROSITE; PS0262; G-PROTEIN RECEPTOR FL 2; 1.
G-protein coupled receptor; Transmembrane.
DOMAIN 1 34 EXTRACELLULAR (POTENTIAL).
TRANSMEM 35 60 1 (POTENTIAL).
DOMAIN 61 64 CYTOPLASMIC (POTENTIAL).
TRANSMEM 65 91 2 (POTENTIAL).
DOMAIN 92 107 EXTRACELLULAR (POTENTIAL).
TRANSMEM 108 129 3 (POTENTIAL).
DOMAIN 130 146 CYTOPLASMIC (POTENTIAL).
TRANSMEM 147 171 4 (POTENTIAL).
DOMAIN 172 197 EXTRACELLULAR (POTENTIAL).
TRANSMEM 198 223 5 (POTENTIAL).
DOMAIN 224 239 CYTOPLASMIC (POTENTIAL).
TRANSMEM 240 264 6 (POTENTIAL).
DOMAIN 265 281 EXTRACELLULAR (POTENTIAL).
TRANSMEM 282 305 7 (POTENTIAL).
DOMAIN 306 355 CYTOPLASMIC (POTENTIAL).
DISULFID 106 183 BY SIMILARITY.
CONFLICT 55 55 M -> V (IN REF. 2 AND 3).
CONFLICT 149 149 L -> F (IN REF. 3).
CONFLICT 278 278 H -> Q (IN REF. 3).
SEQUENCE 355 AA; 40901 MW; FCE9FF70E6F38B1 CRC64;

Query Match 51.5%; Score 51; DB 1; Length 355;
Best Local Similarity 66.7%; Pred. No. 1-2;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AHSVLSFLWTPYAL 17
DB 243 AITLFLWTPYNL 257
| : | | | | |
| : | | | | |

RESULT 2
AURI_SCHPO STANDARD; PRT; 422 AA.
AC Q10142; P79014;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aureobasidin A resistance protein homolog.

```

```

GN AURI OR SPAC3H8.06.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=THR01;
RA Hashida-Okado T.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst A., Thode G.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE AURI FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB000821; BAA19190.1; -.
EMBL; Z69086; CAA93163.1; -.
PIR; T38764; T38764.
GeneDB SPombe; SPAC3H8.06; -.
InterPro; IPR008934; AcPase VanPerase.
InterPro; IPR000326; PA_PTPase.
PFam; PF01569; PAP2; 1.
SMART; SM00014; acidPPC; 1.
Transmembrane.
KW TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 136 156 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT CONFLICT 240 240 G -> C (IN REF. 1).
SQ SEQUENCE 422 AA; 47444 MW; C24EC4E3423E442E CRC64;

Query Match 48.5%; Score 48; DB 1; Length 422;
Best Local Similarity 50.0%; Pred. No. 3.9;

```

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
 142 HSILDLAWPYGV 155

RESULT 3

OPSB_ANOCA STANDARD; PRT; 355 AA.

ID OPSB_ANOCA
 AC P51471;
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Blue-sensitive opsin (Blue photoreceptor pigment) (RH2 opsin).
 OS Anolis carolinensis (Green anole) (American chameleon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Iguania; Iguanidae; Polychrotinae; Anolis.
 OX NCBI_TaxID=28377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95371134; PubMed=7643409;
 RA Kawamura S., Yokoyama S.;
 RT "Paralogous origin of the rhodopsinlike opsin genes in lizards.";
 RL J. Mol. Evol. 40:594-600(1995).
 CC -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL. THIS OPSIN USES A VITAMIN-A2 CHROMOPHORE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- PTM: Some or all of the carboxyl-terminal Ser or Thr residues may
 CC be phosphorylated (By similarity).
 CC -!- MISCELLANEOUS: THIS OPSIN HAS PROBABLY AN ABSORPTION MAXIMUM AT
 CC 503 NM.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC Opsin subfamily.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; S79167; AAB35062.1; JOINED.
 EMBL; S79124; AAB35062.1; JOINED.
 EMBL; S79134; AAB35062.1; JOINED.
 EMBL; S79165; AAB35062.1; JOINED.
 EMBL; S79166; AAB35062.1; JOINED.
 PIR; I51319; I51319.
 HSP; P02699; 1BQJ.
 InterPro; IPR000276; GPCR_Rhodpsn.
 InterPro; IPR001760; Opsin.
 Pfam; PF00001; 7tm.1; 1.
 PRINTS; PR00237; GPCRHHODOPSN.
 PROSITE; PS00237; G PROTEIN RECP F1_1; 1.
 PROSITE; PS50262; G-PROTEIN_RECEP_F1_2; 1.
 PROSITE; PS00238; OPSIN; 1.
 Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 Phosphorylation; G-protein coupled receptor; Lipoprotein; Palmitate.

DOMAIN 1 36
 TRANSMEM 37 61
 DOMAIN 62 73
 TRANSMEM 74 98
 DOMAIN 99 113
 TRANSMEM 114 133
 DOMAIN 134 152
 TRANSMEM 153 176
 DOMAIN 177 202
 TRANSMEM 203 230
 DOMAIN 231 252
 TRANSMEM 253 276
 DOMAIN 277 284

FT TRANSMEM 285 309 7 (POTENTIAL).
 FT DOMAIN 310 355 CYTOPLASMIC.
 FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 110 187 BY SIMILARITY.
 FT BINDING 296 296 RETINAL CHROMOPHORE (BY SIMILARITY).
 FT LIPID 322 322 S-palmitoyl cysteine (By similarity).
 FT LIPID 323 323 S-palmitoyl cysteine (By similarity).
 SQ SEQUENCE 355 AA; 39983 MW; 555B671A4DE9EA6 CRC64;

Query Match 48.0%; Score 47.5; DB 1; Length 355;
 Best Local Similarity 76.9%; Pred. No. 4;
 Matches 10; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 6 VLSFLWTPYAL 17
 258 VLGFLLAWTPYAM 270

Db

RESULT 4

OPS2_HEMSA STANDARD; PRT; 377 AA.

ID OPS2_HEMSA
 AC Q25158;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE Compound eye opsin BCRH2.
 DE Hemigrapsus sanguineus (Crab).
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
 OC Eubrachyura; Grapsoidae; Varunidae; Hemigrapsus.
 OX NCBI_TaxID=40176;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RA Sakamoto K., Hisatomi O., Tokunaga F., Eguchi E.;
 RT "Two opsins from the compound eye of the crab Hemigrapsus
 sanguineus.";
 RL J. Exp. Biol. 199:441-450(1996).
 CC -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL. THIS OPSIN PRODUCES VISUAL PIGMENTS WITH
 CC MAXIMAL ABSORPTION IN THE BLUE-GREEN REGION OF THE SPECTRUM.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN ALL OF THE SEVEN RETINULAR CELLS
 CC (R1-R7) FORMING THE MAIN RHABDOM IN EACH OMATIDIUM.
 CC -!- PTM: Some or all of the carboxyl-terminal Ser or Thr residues may
 CC be phosphorylated (By similarity).
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC Opsin subfamily.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; D50584; BAA09133.1; --
 HSP; P02699; 1F88.
 InterPro; IPR000276; GPCR_Rhodpsn.
 InterPro; IPR001760; Opsin.
 Pfam; PF00001; 7tm.1; 1.
 PRINTS; PR00237; GPCRHHODOPSN.
 PROSITE; PS00237; G PROTEIN RECP F1_1; 1.
 PROSITE; PS50262; G-PROTEIN_RECEP_F1_2; 1.
 PROSITE; PS00238; OPSIN; 1.
 Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 Phosphorylation; G-protein coupled receptor.

DOMAIN 1 53
 TRANSMEM 54 78
 DOMAIN 79 101
 TRANSMEM 102 124
 DOMAIN 125 147
 TRANSMEM 148 170
 DOMAIN 171 193
 TRANSMEM 194 216
 DOMAIN 217 239
 TRANSMEM 240 262
 DOMAIN 263 285
 TRANSMEM 286 308
 DOMAIN 309 331
 TRANSMEM 332 354
 DOMAIN 355 377

FT DOMAIN 79 90 CYTOPLASMIC.
 FT TRANSEM 91 115 2 (POTENTIAL).
 FT DOMAIN 116 131 EXTRACELLULAR.
 FT TRANSEM 132 151 3 (POTENTIAL).
 FT DOMAIN 152 170 CYTOPLASMIC.
 FT TRANSEM 171 190 4 (POTENTIAL).
 FT DOMAIN 195 218 EXTRACELLULAR.
 FT TRANSEM 219 246 5 (POTENTIAL).
 FT DOMAIN 247 281 CYTOPLASMIC.
 FT TRANSEM 282 305 6 (POTENTIAL).
 FT DOMAIN 306 313 EXTRACELLULAR.
 FT TRANSEM 314 338 7 (POTENTIAL).
 FT DOMAIN 339 377 CYTOPLASMIC.
 FT TRANSEM 378 405 8 (POTENTIAL).
 FT BINDING 128 205 BY SIMILARITY.
 FT BINDING 325 325 RETINAL CHROMOPHORE (BY SIMILARITY).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 377 AA; 42114 MW; FD6CC2E0E199A256 CRC64;

Query Match 47.5%; Score 47; DB 1; Length 377;
 Best Local Similarity 69.2%; Pred. No. 5.1;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 LSPFWTPALKS 19
 DB 289 LWFICWTPALIS 301

RESULT 5
 AC2_PANTR
 ID AC2_PANTR STANDARD; PRT; 440 AA.
 AC Q9N2A7;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Muscarinic acetylcholine receptor M2 (Fragment).
 OS CHRM2.
 GN Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate 220;
 RA Kitano T., Kobayakawa H., Saitou N.;
 RT "Silver project";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC ADENYLATE CYCLASE INHIBITION.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB041392; BAA94477.1;
 CC InterPro; IPR000276; GPCR_Rhodopsin.
 CC Pfam; PF00001; 7tm.1; 1.
 CC PRINTS; PR00237; GPCR_Rhodopsin.
 CC PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
 CC PROSITE; PS00262; G-PROTEIN_RECEP_F2_1; 1.
 CC Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 CC Phosphorylation; Multigene family; G-protein coupled receptor.
 CC NON TER 1 1
 CC TRANSEM <1 19 1 (POTENTIAL).
 CC DOMAIN 20 33 CYTOPLASMIC (POTENTIAL).

FT TRANSEM 34 54 2 (POTENTIAL).
 FT DOMAIN 55 71 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 72 93 3 (POTENTIAL).
 FT DOMAIN 94 113 CYTOPLASMIC (POTENTIAL).
 FT TRANSEM 114 136 4 (POTENTIAL).
 FT DOMAIN 137 158 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 159 181 5 (POTENTIAL).
 FT DOMAIN 182 362 CYTOPLASMIC (POTENTIAL).
 FT TRANSEM 363 383 6 (POTENTIAL).
 FT DOMAIN 384 397 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 398 417 7 (POTENTIAL).
 FT DOMAIN 418 440 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 70 150 BY SIMILARITY.
 FT MOD RES 420 420 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 424 424 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 439 439 PHOSPHORYLATION (POTENTIAL).
 SQ SEQUENCE 440 AA; 48854 MW; 12B0324E13D37DDF CRC64;

Query Match 47.5%; Score 47; DB 1; Length 440;
 Best Local Similarity 58.8%; Pred. No. 5.8;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
 DB 118 IAAWVLSFILMAPAIL 134

RESULT 6
 AC2_CHICK
 ID AC2_CHICK STANDARD; PRT; 466 AA.
 AC P30372;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Muscarinic acetylcholine receptor M2.
 OS CHRM2 OR CM2.
 GN Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=91373358; PubMed=1840593;
 RA Tietje K.M., Nathanson N.M.;
 RT "Embryonic chick heart expresses multiple muscarinic acetylcholine
 RT receptor subtypes. Isolation and characterization of a gene encoding
 RT a novel m2 muscarinic acetylcholine receptor with high affinity for
 RT pirenzepine.";
 RL J. Biol. Chem. 266:17382-17387(1991).
 CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC ADENYLATE CYCLASE INHIBITION.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- MISCELLANEOUS: This receptor has a high affinity for pirenzepine.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; MT3217; AAB04106.1;
 CC InterPro; IPR000276; GPCR_Rhodopsin.
 CC Pfam; PF00001; 7tm.1; 1.
 CC PRINTS; PR00237; GPCR_Rhodopsin.
 CC PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
 CC PROSITE; PS00262; G-PROTEIN_RECEP_F2_1; 1.
 CC PROSITE; PS00262; G-PROTEIN_RECEP_F1_2; 1.

KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 KW Phosphorylation; Multigene family; G-protein coupled receptor.
 FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 26 48 1 (POTENTIAL).
 FT DOMAIN 49 62 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 63 83 2 (POTENTIAL).
 FT DOMAIN 84 100 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 101 122 3 (POTENTIAL).
 FT DOMAIN 123 142 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 143 165 4 (POTENTIAL).
 FT DOMAIN 166 187 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 188 210 5 (POTENTIAL).
 FT DOMAIN 211 388 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 389 409 5 (POTENTIAL).
 FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 424 443 7 (POTENTIAL).
 FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 99 179 BY SIMILARITY.
 FT MOD_RES 446 446 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 450 450 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 465 465 PHOSPHORYLATION (POTENTIAL).
 SQ SEQUENCE 466 AA; 51565 MW; 2E3D8241D6168995 CRC64;

Query Match 47.5%; Score 47; DB 1; Length 466;
 Best Local Similarity 58.8%; Pred. No. 6.1;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSLVLSFLWTPTVAL 17
 147 IAAAWLSFILWAPAIL 163

RESULT 7
 AC2 HUMAN STANDARD; PRT; 466 AA.
 ID AC2 HUMAN STANDARD; PRT; 466 AA.
 AC P08172; Q9PIX9;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Muscarinic acetylcholine receptor M2.
 GN CHRM2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97263421; PubMed=3037705;
 RA Bonner T.I., Buckley N.J., Young A.C., Brann M.R.;
 RT "Identification of a family of muscarinic acetylcholine receptor genes.";
 RL Science 237:527-532 (1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=48166632; PubMed=3443095;
 RA Peralta E.G., Ashkenazi A., Winslow J.W., Smith D.H.,
 RA Ramachandran J., Capon D.J.;
 RT "Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors.";
 RL EMO J. 6:3923-3929 (1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RA Publ H.L. III, Ikeda S.R., Atonstam R.S.;
 RT "cDNA clones of human proteins involved in signal transduction sequenced by the Guthrie cDNA resource center (www.cdna.org).";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 27-466 FROM N.A.

RA Kitano T., Kobayakawa H., Saitou N.;
 RT "Silver project.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC ADENYLATE CYCLASE INHIBITION.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M16404; AA51570.1; --
 CC EMBL; X15264; CA33335.1; --
 CC EMBL; AF498916; AA18939.1; --
 CC EMBL; AB041391; BAA94476.1; --
 CC PIR; S10126; S10126.
 CC PDB; 1LUB; 19-JUN-02
 CC Genew; HGNC:1951; CHRM2.
 CC MIM; 118493; --
 CC GO; GO:0005887; C: integral to plasma membrane; TAS.
 CC GO; GO:0004981; F: muscarinic acetylcholine receptor activity; TAS.
 CC GO; GO:0007188; P: G-protein signaling, coupled to GTPase; TAS.
 CC GO; GO:0007207; P: muscarinic acetylcholine receptor, phospho. . .; TAS.
 CC GO; GO:0007399; P: neurogenesis; TAS.
 CC GO; GO:0009405; P: pathogenesis; TAS.
 CC GO; GO:0008016; P: regulation of heart rate; TAS.
 CC InterPro; IPR00276; GPCR_Rhodopsn.
 CC Pfam; PF00001; 7tm.1; 1.
 CC PRINTS; PR00237; GPCR_Rhodopsn
 CC PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
 CC PROSITE; PS0262; G-PROTEIN RECEPTOR FL2; 1.
 CC Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 KW Phosphorylation; Multigene family; G-protein coupled receptor;
 KW 3D-structure.
 FT DOMAIN 1 22 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 23 45 1 (POTENTIAL).
 FT DOMAIN 46 59 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 60 80 2 (POTENTIAL).
 FT DOMAIN 81 97 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 98 119 3 (POTENTIAL).
 FT DOMAIN 120 139 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 140 162 4 (POTENTIAL).
 FT DOMAIN 163 184 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 185 207 5 (POTENTIAL).
 FT DOMAIN 208 388 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 389 409 6 (POTENTIAL).
 FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 424 443 7 (POTENTIAL).
 FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 6 6 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 9 9 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 96 176 BY SIMILARITY.
 FT MOD_RES 446 446 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 450 450 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 465 465 PHOSPHORYLATION (POTENTIAL).
 SQ SEQUENCE 466 AA; 51715 MW; 2FC2FD7748C22BEC CRC64;

Query Match 47.5%; Score 47; DB 1; Length 466;
 Best Local Similarity 58.8%; Pred. No. 6.1;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSLVLSFLWTPTVAL 17
 :|||:|||||


```

FT TRANSMEM 185 207 5 (POTENTIAL).
FT DOMAIN 208 388 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 389 409 6 (POTENTIAL).
FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 424 443 7 (POTENTIAL).
FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 2 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 6 6 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 96 176 BY SIMILARITY.
FT MOD RES 446 446 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 450 450 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 465 465 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 330 330 K -> N (IN REF. 2).
FT SEQUENCE 466 AA; 51673 MW; 53D089F179789CD9 CRC64;

Query Match 47.5%; Score 47; DB 1; Length 466;
Best Local Similarity 58.8%; Pred. No. 6.1;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYAL 17
Db 144 IAAWVLSFILWAPAIL 160

RESULT 10
ACM2 RAT STANDARD; PRT; 466 AA.
AC P10980; Q92221;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Muscarinic acetylcholine receptor M2.
GN CHR2 OR CHR2-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88068581; PubMed=2825184;
RA Gocayne J.D., Robinson D.A., Fitzgerald M.G., Chung F.-Z.,
RA Kerlavage A.R., Lentes K.-U., Lai J., Wang C.-D., Fraser C.M.,
RA Venter J.C.;
RT "Primary structure of rat cardiac beta-adrenergic and muscarinic
RT cholinergic receptors obtained by automated DNA sequence analysis:
RT further evidence for a multigene family.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8296-8300(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Iris;
RM MEDLINE=99138467; PubMed=9972520;
RA Furuta M., Ohya S., Imaizumi Y., Watanabe M.;
RT "Molecular cloning of m3 muscarinic acetylcholine receptor in rat
RT iris.";
RN [3]
RP J. Smooth Muscle Res. 34:111-122(1998).
CC -1- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
CC BREAKDOWN OF PHOSPHOLIPIDS & MODULATION OF POTASSIUM CHANNELS
CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
CC ADENYLATE CYCLASE INHIBITION.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; J03025; AAA40926.1; -.

```

```

DR EMBL; AB017655; BAA36838.1; -.
DR PIR; S10856; S10856.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECP_F2_1; 1.
KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
KW Phosphorylation; Multigene family; G-protein coupled receptor.
FT DOMAIN 1 22 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 23 45 1 (POTENTIAL).
FT DOMAIN 46 59 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 60 80 2 (POTENTIAL).
FT DOMAIN 81 97 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 98 119 3 (POTENTIAL).
FT DOMAIN 120 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 162 4 (POTENTIAL).
FT DOMAIN 163 184 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 185 207 5 (POTENTIAL).
FT DOMAIN 208 388 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 389 409 6 (POTENTIAL).
FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 424 443 7 (POTENTIAL).
FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 6 176 BY SIMILARITY.
FT MOD RES 446 446 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 450 450 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 465 465 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 51 51 N -> S (IN REF. 1).
FT CONFLICT 273 273 N -> T (IN REF. 1).
FT CONFLICT 289 290 VS -> SA (IN REF. 1).
FT CONFLICT 313 313 G -> D (IN REF. 1).
FT CONFLICT 337 337 C -> Y (IN REF. 1).
FT CONFLICT 353 353 N -> S (IN REF. 1).
FT CONFLICT 360 360 I -> V (IN REF. 1).
FT CONFLICT 369 369 T -> P (IN REF. 1).
FT SEQUENCE 466 AA; 51539 MW; 70ECCD86366A676B CRC64;

Query Match 47.5%; Score 47; DB 1; Length 466;
Best Local Similarity 58.8%; Pred. No. 6.1;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYAL 17
Db 144 IAAWVLSFILWAPAIL 160

RESULT 11
ACM4 RAT STANDARD; PRT; 478 AA.
AC P08485;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Muscarinic acetylcholine receptor M4.
GN CHR4 OR CHR4-4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87263421; PubMed=3037705;
RA Bonner T.I., Buckley N.J., Young A.C., Brann M.R.;
RT "Identification of a family of muscarinic acetylcholine receptor
RT genes.";
RL Science 237:527-532(1987).
CC -1- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE.
CC BREAKDOWN OF PHOSPHOLIPIDS & MODULATION OF POTASSIUM CHANNELS
CC -----
CC EMBL; J03025; AAA40926.1; -.

```

THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 INHIBITION OF ADENYLATE CYCLASE
 -1- SUBCELLULAR LOCATION: Integral membrane protein.
 -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL: M16405; AAA51571.1; ALT SEQ.
 InterPro: IPR000276; GPCR_Rhodpsn.
 PRINTS: PR00237; GPCRHOPOPSN.
 PROSITE: PS00237; G PROTEIN RECEPTOR FL 1; 1.
 PROSITE: PS0262; G PROTEIN RECEPTOR FL 2; 1.
 Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 Phosphorylation; Multigene family; G-protein coupled receptor.
 DOMAIN 1 30 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 31 53 1 (POTENTIAL).
 DOMAIN 54 67 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 68 88 2 (POTENTIAL).
 DOMAIN 89 105 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 106 127 3 (POTENTIAL).
 DOMAIN 128 147 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 148 170 4 (POTENTIAL).
 DOMAIN 171 192 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 193 215 5 (POTENTIAL).
 DOMAIN 216 400 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 401 421 6 (POTENTIAL).
 DOMAIN 422 435 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 436 455 7 (POTENTIAL).
 DOMAIN 456 478 CYTOPLASMIC (POTENTIAL).
 CARBOHYD 13 13 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DISULFID 104 184 BY SIMILARITY.
 MOD_RES 458 458 PHOSPHORYLATION (POTENTIAL).
 MOD_RES 462 462 PHOSPHORYLATION (POTENTIAL).
 MOD_RES 476 476 PHOSPHORYLATION (POTENTIAL).
 SEQUENCE 478 AA; 52920 MW; 2551FF5F7B1A56C CRC64;
 Query Match 46.5%; Score 46; DB 1; Length 478;
 Best Local Similarity 58.8%; Pred. No. 9;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSVLSFLWTPYAL 17
 :|||:||||:|
 Db 152 IAAAVLSFVLWAPAIL 168
 RESULT 12
 ACW4 HUMAN STANDARD; PRT; 479 AA.
 ID ACW4 HUMAN STANDARD; PRT; 479 AA.
 AC P08173;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Muscarinic acetylcholine receptor M4.
 GN CHRM4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87263421; PubMed=3037705;
 RA Bonner T.I., Buckley N.J., Young A.C., Brann M.R.;
 RA "Identification of a family of muscarinic acetylcholine receptor
 genes.";
 RT Science 237:527-532(1987).

RN REVISIONS TO 258 AND 261-266.
 RP MEDLINE=90166521; PubMed=3272174;
 RX Bonner T.I., Young A.C., Brann M.R., Buckley N.J.;
 RA "Cloning and expression of the human and rat m5 muscarinic
 acetylcholine receptor genes.";
 RL Neuron 1:403-410(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88166632; PubMed=3443095;
 RA Peralta E.G., Ashkenazi A., Winslow J.W., Smith D.H.,
 RA Ramachandran J., Capon D.J.;
 RT "Distinct primary structures, ligand-binding properties and tissue-
 specific expression of four human muscarinic acetylcholine
 receptors.";
 RL EMBO J. 6:3923-3929(1987).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RC Puhl H.L. III, Ikeda S.R., Aronstam R.S.;
 RA "cDNA clones of human proteins involved in signal transduction
 sequenced by the Guthrie cDNA resource center (www.cdna.org).";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: The muscarinic acetylcholine receptor mediates various
 cellular responses, including inhibition of adenylate cyclase,
 breakdown of phosphoinositides and modulation of potassium
 channels through the action of G proteins. Primary transducing
 effect is inhibition of adenylate cyclase.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL: M16405; AAA51571.1; ALT_SEQ.
 EMBL: X15285; CA33336.1; --
 EMBL: AF498918; AAM18941.1; --
 PIR: S10127; S10127.
 DR HSSP: P02699; 1BOJ.
 DR MIM: 118495; --
 DR GO: GO:0005887; C-integral to plasma membrane; TAS.
 DR GO: GO:0004981; F-muscarinic acetylcholine receptor activity; TAS.
 DR GO: GO:0007213; P-acetyl choline receptor signaling, muscarin. . .; TAS.
 DR GO: GO:0008283; P-cell proliferation; TAS.
 DR GO: GO:0007166; P-cell surface receptor linked signal transdu. . .; TAS.
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PRINTS: PR00237; GPCRHOPOPSN.
 DR PROSITE: PS00237; G PROTEIN RECEPTOR FL 1; 1.
 DR PROSITE: PS0262; G PROTEIN RECEPTOR FL 2; 1.
 KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 KW Phosphorylation; Multigene family; G-protein coupled receptor.
 DOMAIN 1 31 1 (POTENTIAL).
 TRANSMEM 32 54 1 (POTENTIAL).
 DOMAIN 55 68 2 (POTENTIAL).
 TRANSMEM 69 89 2 (POTENTIAL).
 DOMAIN 90 106 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 107 128 3 (POTENTIAL).
 DOMAIN 129 148 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 149 171 4 (POTENTIAL).
 DOMAIN 172 193 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 194 216 5 (POTENTIAL).
 DOMAIN 217 401 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 402 422 6 (POTENTIAL).
 DOMAIN 423 436 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 437 456 7 (POTENTIAL).
 DOMAIN 457 479 CYTOPLASMIC (POTENTIAL).

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 13 13 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 105 185 BY SIMILARITY.
 FT MOD RES 459 459 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 463 463 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 477 477 PHOSPHORYLATION (POTENTIAL).
 FT CONFLICT 184 184 Q -> H (IN REF. 1).
 SQ SEQUENCE 479 AA; 53049 MW; 9CIB05B359A64289 CRC64;
 Query Match 46.5%; Score 46; DB 1; Length 479;
 Best Local Similarity 58.8%; Pred. No. 9;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSLVSLFWLWTPYAL 17
 Db 153 IAAAWLSFVLWAPAIL 169
 RESULT 13
 ID ACMA_MOUSE STANDARD; PRT; 479 AA.
 AC P32211; Q64036;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Muscarinic acetylcholine receptor M4 (Mm4 mAChR).
 GN CHRM4 OR CHRM-4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 1
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93305731; PubMed=7916637;
 RA van Koppen C.J., Lenz W., Nathanson N.M.;
 RT "Isolation, sequence and functional expression of the mouse m4
 RT muscarinic acetylcholine receptor gene";
 RL Biochim. Biophys. Acta 1173:342-344 (1993).
 RN [2]
 RP SEQUENCE OF 235-355 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95179320; PubMed=7874308;
 RX Andre C., Dos Santos G., Koulakoff A.;
 RA "Cultured neurons from mouse brain reproduce the muscarinic receptor
 RT profile of their tissue of origin.";
 RL Eur. J. Neurosci. 6:1691-1701 (1994).
 CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC INHIBITION OF ADENYLATE CYCLASE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X63473; CAA45071.1; -;
 CC EMBL; S74916; AAB33577.1; -;
 CC PIR; S33776; S33776.
 CC HSSP; P02699; 1BOJ.
 CC MGD; MGI:88399; Chrm4.
 CC InterPro; IPR000276; GPCR_Rhodopsn.
 CC Pfam; PF00001; 7cm1; 1.
 CC PRINTS; PR00237; GPCRHHODOPSN.
 CC PROSITE; PS00237; G PROTEIN RECEPTOR FL 1; 1.
 CC PROSITE; PS0262; G PROTEIN RECEPTOR FL 2; 1.
 CC Postsynaptic membrane; Ionc channel; Glycoprotein; Transmembrane;
 KW CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW CARBOHYD 13 13 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW DISULFID 105 185 BY SIMILARITY.
 KW MOD RES 459 459 PHOSPHORYLATION (POTENTIAL).
 KW MOD RES 463 463 PHOSPHORYLATION (POTENTIAL).
 KW MOD RES 477 477 PHOSPHORYLATION (POTENTIAL).
 KW CONFLICT 184 184 Q -> H (IN REF. 1).
 SQ SEQUENCE 479 AA; 53049 MW; 9CIB05B359A64289 CRC64;
 Query Match 46.5%; Score 46; DB 1; Length 479;
 Best Local Similarity 58.8%; Pred. No. 9;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSLVSLFWLWTPYAL 17
 Db 153 IAAAWLSFVLWAPAIL 169
 RESULT 13
 ID ACMA_MOUSE STANDARD; PRT; 479 AA.
 AC P32211; Q64036;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Muscarinic acetylcholine receptor M4 (Mm4 mAChR).
 GN CHRM4 OR CHRM-4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 1
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93305731; PubMed=7916637;
 RA van Koppen C.J., Lenz W., Nathanson N.M.;
 RT "Isolation, sequence and functional expression of the mouse m4
 RT muscarinic acetylcholine receptor gene";
 RL Biochim. Biophys. Acta 1173:342-344 (1993).
 RN [2]
 RP SEQUENCE OF 235-355 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95179320; PubMed=7874308;
 RX Andre C., Dos Santos G., Koulakoff A.;
 RA "Cultured neurons from mouse brain reproduce the muscarinic receptor
 RT profile of their tissue of origin.";
 RL Eur. J. Neurosci. 6:1691-1701 (1994).
 CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC INHIBITION OF ADENYLATE CYCLASE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X63473; CAA45071.1; -;
 CC EMBL; S74916; AAB33577.1; -;
 CC PIR; S33776; S33776.
 CC HSSP; P02699; 1BOJ.
 CC MGD; MGI:88399; Chrm4.
 CC InterPro; IPR000276; GPCR_Rhodopsn.
 CC Pfam; PF00001; 7cm1; 1.
 CC PRINTS; PR00237; GPCRHHODOPSN.
 CC PROSITE; PS00237; G PROTEIN RECEPTOR FL 1; 1.
 CC PROSITE; PS0262; G PROTEIN RECEPTOR FL 2; 1.
 CC Postsynaptic membrane; Ionc channel; Glycoprotein; Transmembrane;
 KW CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW CARBOHYD 13 13 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW DISULFID 104 184 BY SIMILARITY.
 KW MOD RES 459 459 PHOSPHORYLATION (POTENTIAL).
 KW MOD RES 463 463 PHOSPHORYLATION (POTENTIAL).
 KW CONFLICT 322 323 AL -> V (IN REF. 2).
 KW CONFLICT 351 351 S -> N (IN REF. 2).
 SQ SEQUENCE 479 AA; 52973 MW; 04F301E78814CD70 CRC64;
 Query Match 46.5%; Score 46; DB 1; Length 479;
 Best Local Similarity 58.8%; Pred. No. 9;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSLVSLFWLWTPYAL 17
 Db 152 IAAAWLSFVLWAPAIL 168
 RESULT 14
 ID ACMA_CHICK STANDARD; PRT; 490 AA.
 AC P17200;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Muscarinic acetylcholine receptor M4.
 GN CHRM4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90153912; PubMed=2154460;
 RX Tietje K.M., Goldman P.S., Nathanson N.M.;
 RA "Cloning and functional analysis of a gene encoding a novel
 RT muscarinic acetylcholine receptor expressed in chick heart and
 RT brain.";
 RL J. Biol. Chem. 265:2828-2834 (1990).
 CC -!- FUNCTION: The muscarinic acetylcholine receptor mediates various
 CC cellular responses, including inhibition of adenylate cyclase,
 CC breakdown of phosphoinositides and modulation of potassium
 CC channels through the action of G proteins. Primary transducing
 CC effect is inhibition of adenylate cyclase. May couple to multiple
 CC functional responses in cell lines.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed in heart and brain.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

or send an email to license@sib-sib.ch.

```

CC CC EMBL; J05218; AAA48563.1; -.
CC CC PIR; A35546; A35546.
CC CC HSP; P02699; 1BOU.
CC CC InterPro; IPR000276; GPCR Rhodpsn.
CC CC Pfam; PF00001; 7tm_1; 1.
CC CC PRINTS; PR00237; GPCRHOPOPSN.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_1; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_2; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_3; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_4; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_5; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_6; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_7; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_8; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_9; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_10; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_11; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_12; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_13; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_14; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_15; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_16; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_17; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_18; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_19; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_20; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_21; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_22; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_23; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_24; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_25; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_26; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_27; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_28; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_29; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_30; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_31; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_32; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_33; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_34; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_35; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_36; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_37; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_38; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_39; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_40; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_41; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_42; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_43; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_44; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_45; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_46; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_47; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_48; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_49; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_50; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_51; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_52; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_53; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_54; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_55; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_56; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_57; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_58; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_59; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_60; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_61; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_62; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_63; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_64; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_65; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_66; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_67; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_68; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_69; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_70; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_71; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_72; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_73; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_74; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_75; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_76; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_77; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_78; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_79; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_80; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_81; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_82; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_83; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_84; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_85; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_86; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_87; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_88; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_89; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_90; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_91; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_92; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_93; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_94; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_95; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_96; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_97; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_98; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_99; 1.
CC CC PROSITE; PS00237; G PROTEIN RECF Fl_100; 1.

```

Query Match 46.5%; Score 46; DB 1; Length 490;

Best Local Similarity 52.9%; Pred. No. 9.2;

Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 MAHVSLSFLWTPYAL 17
Db 163 IAAWLSFILWAPAIL 179

```

RESULT 15

```

ID OPSB GECE STANDARD; PRT; 355 AA.
AC P35357;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Blue-sensitive opsin P467 (Blue photoreceptor pigment).
OS Gecko gecko (Tokay gecko).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Gekkota; Gekkonidae; Gekko.
OX NCBI_TaxID=36310;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92357726; PubMed=1379723;
RA Kojima D., Okano T., Fukada Y., Shichida Y., Yoshizawa T.,
RA Ebrey T.G.;
RT "Cone visual pigments are present in gecko rod cells."
RL Proc. Natl. Acad. Sci. U.S.A. 89:6841-6845(1992).
CC -!- FUNCTION: Visual pigments are the light-absorbing molecules that
CC mediate vision. They consist of an apoprotein, opsin, covalently
CC linked to cis-retinal.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: IN THIS LIZARD THE COLOR PIGMENTS ARE FOUND IN
CC THE ROD-SHAPED PHOTORECEPTOR CELLS WHICH HAVE BEEN DERIVED FROM
CC ANCESTRAL CONE-LIKE PHOTORECEPTORS.
CC -!- PTM: Some or all of the carboxyl-terminal Ser or Thr residues may
CC be phosphorylated (By similarity).
CC -!- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMUM AT 467 NM.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.

```

Opsin subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

```

CC EMBL; M20205; AAA49307.1; -.
CC PIR; A46191; A46191.
CC HSP; P02699; 1EDX.
CC InterPro; IPR000276; GPCR Rhodpsn.
CC Pfam; PF00001; 7tm_1; 1.
CC PRINTS; PR00237; GPCRHOPOPSN.
CC PROSITE; PS00237; G PROTEIN RECF Fl_1; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_2; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_3; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_4; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_5; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_6; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_7; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_8; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_9; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_10; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_11; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_12; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_13; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_14; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_15; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_16; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_17; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_18; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_19; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_20; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_21; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_22; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_23; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_24; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_25; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_26; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_27; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_28; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_29; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_30; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_31; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_32; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_33; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_34; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_35; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_36; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_37; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_38; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_39; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_40; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_41; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_42; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_43; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_44; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_45; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_46; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_47; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_48; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_49; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_50; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_51; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_52; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_53; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_54; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_55; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_56; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_57; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_58; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_59; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_60; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_61; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_62; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_63; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_64; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_65; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_66; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_67; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_68; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_69; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_70; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_71; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_72; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_73; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_74; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_75; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_76; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_77; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_78; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_79; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_80; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_81; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_82; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_83; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_84; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_85; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_86; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_87; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_88; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_89; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_90; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_91; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_92; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_93; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_94; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_95; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_96; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_97; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_98; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_99; 1.
CC PROSITE; PS00237; G PROTEIN RECF Fl_100; 1.

```

Query Match 46.0%; Score 45.5; DB 1; Length 355;
 Best Local Similarity 83.3%; Pred. No. 8.2;
 Matches 10; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

QY 6 VLSFLL-WTPYA 16
Db 258 VLGFLLAWTPYA 269

```

Search completed: July 19, 2004, 17:05:41
 Job time : 13 secs

BAC IG005110.
GN A.IG005110.24 OR F5110.24 OR AT4G00450.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Andrews S.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Waterston R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Wilson R., Andrews S.;
RL "The sequence of A. thaliana F5110.";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA WashU;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Waterston R.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF013293; AAB62842.1; -;
DR EMBL; AF195115; AAF02800.1; -;
DR EMBL; AL161472; CAB80854.1; -;
DR PIR; T01526; T01526.
KW Hypothetical protein.
SQ SEQUENCE 2124 AA; 235154 MW; 84892F6BEF3B65D5 CRC64;
Query Match 49.5%; Score 49; DB 10; Length 2124;
Best Local Similarity 41.2%; Pred. No. 26;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 3 AHSVLSFLWTPTALKS 19
Db 1438 AYSIIAFVLRPFVSNS 1454
RESULT 6
Q9N2A5 PRELIMINARY; PRT; 440 AA.
AC Q9N2A5;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Muscarinic acetylcholine receptor m2 (Fragment).
GN CHRM2.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=oran-UL;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AB041394; BAA94479.1; -;
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G-PROTEIN RECP_F1_1; 1.
DR PROSITE; PS0262; G-PROTEIN RECP_F1_2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 440 AA; 48870 MW; BDB2FEB1AC3B83E CRC64;
Query Match 47.5%; Score 47; DB 6; Length 440;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSVLSFLWTPTVAL 17
Db 118 IAAAWLSFILWAPAIL 134
RESULT 7
Q9N2A7 PRELIMINARY; PRT; 440 AA.
AC Q9N2A7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Muscarinic acetylcholine receptor m2 (Fragment).
GN CHRM2.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=chimp-220;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AB041392; BAA94477.1; -;
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G-PROTEIN RECP_F1_1; 1.
DR PROSITE; PS0262; G-PROTEIN RECP_F1_2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 440 AA; 48854 MW; 12B0324E13D37DDF CRC64;
Query Match 47.5%; Score 47; DB 6; Length 440;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSVLSFLWTPTVAL 17
Db 118 IAAAWLSFILWAPAIL 134

```

RESULT 8
Q9N2A6 PRELIMINARY; PRT; 440 AA.
ID Q9N2A6
AC Q9N2A6; 2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 24, Last annotation update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscarinic acetylcholine receptor m2 (Fragment).
GN CHRM2.
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=gorilla-UL;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AB041393; BA94478.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL 2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON TER 1
SQ SEQUENCE 440 AA; 48954 MW; 12B0324E13D37DDF CRC64;

Query Match 47.5%; Score 47; DB 6; Length 440;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
Db 118 IAAAVLSFILWAPAIL 134
: || |||||: ||

RESULT 9
Q96RH0 PRELIMINARY; PRT; 456 AA.
ID Q96RH0
AC Q96RH0; 2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE M2 muscarinic cholinergic receptor.
GN CHRM2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Aredondo J., Grando S.A.;
RT "Cloning Cholinergic Receptors in Human Keratinocytes.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AF385588; AA68113.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.

RESULT 10
Q8VH27 PRELIMINARY; PRT; 466 AA.
ID Q8VH27
AC Q8VH27; 2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscarinic receptor 2 (Fragment).
GN GPM2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RA So I., Yang D., Kim H., Min K., Kim S., Kim K., Park K., Choi K.,
Kim I.;
RT "Five subtypes of muscarinic receptors are expressed in gastric smooth
muscles of guinea pig.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AY072059; AAL67910.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL 2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON TER 466
SQ SEQUENCE 466 AA; 51630 MW; A992CB6766A37BD5 CRC64;

Query Match 47.5%; Score 47; DB 11; Length 466;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
Db 144 IAAAVLSFILWAPAIL 160
: || |||||: ||

RESULT 11
Q80LM4 PRELIMINARY; PRT; 495 AA.
ID Q80LM4
AC Q80LM4; 2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 25, Last annotation update)
DE M2 muscarinic receptor.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]

```

```

RESULT 13
O88616 PRELIMINARY; PRT; 166 AA.
ID AC O88616;
DT 01-NOV-1998 (TtEMBLrel. 08, Created)
DT 01-NOV-1998 (TtEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TtEMBLrel. 24, Last annotation update)
DE Muscarinic acetylcholine receptor M4 (fragment).
OS Meriones.
OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Meriones.
OC NCBI_TaxID=10047;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21431793; PubMed=11547340;
RA Wangemann P., Liu J., Scherz E.Q., Herzog M., Shimozone M.,
RA Scofield M.A.;
RT "Muscarinic receptors control K+ secretion in inner ear strial
RT marginal cells.";
RL J. Membr. Biol. 182:171-181(2001).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AF079114; AAC28098.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:00041584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodopsin.
DR Pfam; PF00001; 7tm1; 1.
DR PRINTS; PR00237; GPCR_RHODOPSIN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON_TER 1
FT NON_TER 166
FT NON_TER 166
SQ SEQUENCE 166 AA; 18477 MW; B13F1DC7B73DF41F CRC64;

Query Match 46.5%; Score 46; DB 11; Length 166;
Best Local Similarity 58.8%; Pred. No. 6.5;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps

QY 1 MAASHVLSFLLWTPYAL 17
: || ||||| : |
DB 120 IAAAVLSFVLMAPAIL 136

RESULT 14
Q96RG8 PRELIMINARY; PRT; 474 AA.
ID AC Q96RG8;
DT 01-DEC-2001 (TtEMBLrel. 19, Created)
DT 01-DEC-2001 (TtEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TtEMBLrel. 24, Last annotation update)
DE M4 muscarinic cholinergic receptor.
OS ChRM4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A.
RC Arredondo J., Grando S.A.;
RT "Cloning Cholinergic Receptors in Human Keratinocytes.";
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AF385590; AAK68115.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.

```

DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodopsn.

DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 474 AA; 52507 MW; 061B59A48A3FEFC8 CRC64;

Query Match 46.5%; Score 46; DB 4; Length 474;
Best Local Similarity 58.8%; Pred.No.19;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSHVLSFLWTPYAL 17
:|||||:|
Db 153 IAAAWLSFVIMAPAIL 169

RESULT 15

Q8VH25 PRELIMINARY; PRT; 477 AA.
AC Q8VH25;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscarinic receptor 4 (Fragment).
GN GPM4.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RA So I., Yang D., Kim H., Min K., Kim S., Kim K., Park K., Choi K.,
RA Kim I.;
RT "Five subtypes of muscarinic receptors are expressed in gastric smooth
RT muscles of guinea pig."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AY072061; AL67912.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON TER 477 477
SQ SEQUENCE 477 AA; 52767 MW; 8B2E19B2F4EABC6A CRC64;

Query Match 46.5%; Score 46; DB 11; Length 477;
Best Local Similarity 58.8%; Pred.No.19;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSHVLSFLWTPYAL 17
:|||||:|
Db 152 IAAAWLSFVIMAPAIL 168

Search completed: July 19, 2004, 17:06:30
Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:07:01 ; Search time 54 Seconds
(without alignments)
104.647 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHVSLSFLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	95.0	19	6	ADA40708 Human sec
2	19	95.0	19	6	ABR47713 Human sec
3	19	95.0	20	2	AY07853 Human sec
4	7	35.0	57	5	ABP32031 Human ORF
5	7	35.0	113	4	AAM83410 Human imm
6	7	35.0	285	4	AB79589 Coryneb
7	7	35.0	355	2	AAW29179 Rat CC ch
8	7	35.0	355	7	ADD45360 Rat Prote
9	7	35.0	1019	4	ABG92745 C glutami
10	7	35.0	1019	5	AB79444 Coryneb
11	6	30.0	33	2	AY60093 Human end
12	6	30.0	33	4	AB64900 Human sec
13	6	30.0	33	4	AB64424 Human sec
14	6	30.0	38	3	ABG24890 Human sec
15	6	30.0	41	3	AB44612 Human sec
16	6	30.0	41	4	AAM06857 Human foe
17	6	30.0	41	4	AAM06447 Human foe
18	6	30.0	55	5	ABP00846 Human ORF
19	6	30.0	59	2	AAE68891 Japanese
20	6	30.0	59	3	ABG58213 Arabidops
21	6	30.0	62	5	ABP35008 Human ORF
22	6	30.0	63	3	AB23813 Arabidops
23	6	30.0	63	6	ABP99744 Human sec
24	6	30.0	63	6	ABR01235 Human gen
25	6	30.0	63	7	ADC20501 Human sec

ALIGNMENTS

RESULT 1

ADA40708
ID ADA40708 standard; protein; 19 AA.

XX AC ADA40708;

XX DT 20-NOV-2003 (first entry)

XX DE Human secreted protein.

XX KW Human; secreted protein; cancer; hyperproliferative disorder;

KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;

KW anaemia; allergic reaction; asthma; cardiovascular disorder;

KW wound healing; cytostatic; immunosuppressive; neutropenic; neuroprotective;

KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;

KW vulnary; cardiant; gene therapy.

XX OS Homo sapiens.

XX PN WO2002102993-A2.

XX PD 27-DEC-2002.

XX PF 19-MAR-2002; 2002WO-US008123.

XX PR 21-MAR-2001; 2001US-0277340P.

XX PR 19-JUL-2001; 2001US-0306171P.

XX PR 13-NOV-2001; 2001US-0331287P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI; 2003-175238/17.

XX PT New human secreted proteins and nucleic acid molecules, useful for

XX PT preparing a diagnostic or pharmaceutical composition for diagnosing,

XX PT preventing or treating cancer or other hyperproliferative disorder,

XX PT asthma, allergies or AIDS.

XX PS Claim 1; SEQ ID NO 1090; 3205pp; English.

XX CC The invention relates to novel genes ADA39629-ADA40565 and proteins

XX CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,

XX CC treating or ameliorating medical conditions e.g. by protein or gene

XX CC therapy. The polypeptides, nucleic acid molecules, antibodies or their

XX CC fragments, and agonists or antagonists that bind to the polypeptide are

XX CC useful for preparing a diagnostic or pharmaceutical composition for

Aam99722 Human exc
Aam42537 Human kid
Ada40922 Human sec
Adc74224 Human sec
Aay02749 Human sec
Ada07428 Human sec
Aam81705 Human bae
Aam81448 Human bae
Aam00848 Human bon
Aam00848 Human bon
Aau51440 Propionib
Aba47959 Propionib
Aaol11233 Human pol
Aaol11233 Human pol
Aam94382 Human rep
Aau65663 Propionib
Aam62182 Propionib
Aay87138 Human sec
Aae06115 Human gen
Abg333937 Human sec
Abp07585 Human ORF
Aaol11478 Human pol

CC diagnosing or treating cancer or other hyperproliferative disorder. The
 CC polypeptides and nucleic acid molecules are also useful for detecting,
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
 CC or other hyperproliferative disorders including neoplasms, autoimmune
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
 CC thrombocytopenia), allergic reactions including asthma or eczema,
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
 CC fungal or viral infections including HIV/AIDS), or wound healing and
 CC disorders of epithelial cell proliferation. The nucleic acids are also
 CC useful for chromosome identification, radiation hybrid mapping or long-
 CC range restriction mapping, as molecular weight markers, or as
 CC hybridization or diagnostic probes. The polypeptides and antibodies are
 CC useful for providing immunological probes for differential identification
 CC of the tissues immunohistochemistry assays. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 19 AA;

Query Match 95.0%; Score 19; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLWTPYALKS 19
 |||||
 Db 1 MAAHSVLSFLWTPYALKS 19

RESULT 2

ABR47713
 ID ABR47713 standard; protein; 19 AA.

AC ABR47713;

DT 12-JUN-2003 (first entry)

DE Human secreted protein, SEQ ID 604.

XX Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
 KW vulnaray; antiinflammatory; nootropic; neuroprotective;
 KW antiparkinsonian; gene therapy; human; cardiovascular disorder.

XX Homo sapiens.

OS WO200295010-A2.

PN 28-NOV-2002.

XX 19-MAR-2002; 2002WO-US009785.

XX 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-129429/12.

XX Novel human secreted proteins, useful for detecting, preventing,
 PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
 PT disorders such as arrhythmia.

XX Claim 13; SEQ ID NO 604; 188pp; English.

XX The present invention relates to novel human secreted proteins (ABR47633-

CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
 CC and their coding sequences are useful for the preparation of a diagnostic
 CC or pharmaceutical composition for diagnosing or treating a cardiovascular
 CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
 CC arteriosclerosis and myocardial ischaemia), neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, renal disorders,
 CC proliferative disorders and/or cancerous diseases and conditions, for
 CC wound healing and epithelial cell proliferation, to treat inflammation or
 CC infection, for treating thrombosis and arteriosclerosis, for treating or
 CC preventing neural damage which occurs in neuronal disorders or
 CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
 CC disease, to enhance bone and periodontal regeneration and aid in tissue
 CC transplants or bone grafts, to prevent skin aging or hair loss, to
 CC stimulate growth and differentiation of haematopoietic cells and bone
 CC marrow cells when used in combination with other cytokines, to maintain
 CC organs before transplantation or for supporting cell culture of primary
 CC tissues, to increase or decrease differentiation or proliferation of
 CC embryonic stem cells, or to modulate mammalian characteristics or
 CC metabolism. Note: The sequence data for this patent was published in
 CC electronic format and is available from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 19 AA;

Query Match 95.0%; Score 19; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLWTPYALKS 19
 |||||
 Db 1 MAAHSVLSFLWTPYALKS 19

RESULT 3

AAV07853
 ID AAV07853 standard; protein; 20 AA.

AC AAV07853;

DT 06-JUL-1999 (first entry)

DE Human secreted protein fragment encoded from gene 2.

XX Human; secreted protein; treatment; prevention; protein therapy; AIDS;
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
 KW arthritis; malignancy; digestive; infection.

OS Homo sapiens.

XX WO9918208-A1.

XX 15-APR-1999.

XX 01-OCT-1998; 98WO-US020775.

XX 02-OCT-1997; 97US-0060833P.

XX 02-OCT-1997; 97US-0060836P.

XX 02-OCT-1997; 97US-0060837P.

XX 02-OCT-1997; 97US-0060838P.

XX 02-OCT-1997; 97US-0060839P.

XX 02-OCT-1997; 97US-0060843P.

XX 02-OCT-1997; 97US-0060862P.

XX 02-OCT-1997; 97US-0060866P.

XX 02-OCT-1997; 97US-0060874P.

XX 02-OCT-1997; 97US-0060880P.

XX 02-OCT-1997; 97US-0060884P.

DE Human immune/haematopoietic antigen SEQ ID NO:11003.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
OS Homo sapiens.
XX WO200157182-A2.
PN 09-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001354.
PF
XX
31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184564P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0230348P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-024617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251989P.

PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-483426/52.
 DR N-PSDB; AAK56191.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS Claim 11; SEQ ID NO 11003; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 113 AA;

Query Match 35.0%; Score 7; DB 4; Length 113;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFLL 11
 Db 34 SVLSFLL 40

RESULT 6
 AAB79589
 ID AAB79589 standard; protein; 285 AA.
 XX
 AC AAB79589;
 XX
 XX 30-APR-2001 (first entry)
 DT
 DE
 XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:694.
 XX
 KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
 XX
 OS Corynebacterium glutamicum.
 XX
 XX WO200100844-A2.
 XX
 XX 04-JAN-2001.
 XX
 XX 23-JUN-2000; 2000WO-IB000943.
 XX
 XX 25-JUN-1999; 99US-0141031P.
 PR 08-JUL-1999; 99DE-01031412.

PR 08-JUL-1999; 99DE-01031413.
 PR 08-JUL-1999; 99DE-01031419.
 PR 08-JUL-1999; 99DE-01031420.
 PR 08-JUL-1999; 99DE-01031424.
 PR 08-JUL-1999; 99DE-01031428.
 PR 08-JUL-1999; 99DE-01031431.
 PR 08-JUL-1999; 99DE-01031433.
 PR 08-JUL-1999; 99DE-01031434.
 PR 08-JUL-1999; 99DE-01031510.
 PR 08-JUL-1999; 99DE-01031562.
 PR 08-JUL-1999; 99DE-01031634.
 PR 09-JUL-1999; 99DE-01032180.
 PR 09-JUL-1999; 99DE-01032227.
 PR 09-JUL-1999; 99DE-01032230.
 PR 09-JUL-1999; 99US-0143208P.
 PR 14-JUL-1999; 99DE-01032924.
 PR 14-JUL-1999; 99DE-01032973.
 PR 14-JUL-1999; 99DE-01033005.
 PR 27-AUG-1999; 99DE-01040765.
 PR 31-AUG-1999; 99US-0151572P.
 PR 03-SEP-1999; 99DE-01042076.
 PR 03-SEP-1999; 99DE-01042079.
 PR 03-SEP-1999; 99DE-01042086.
 PR 03-SEP-1999; 99DE-01042087.
 PR 03-SEP-1999; 99DE-01042088.
 PR 03-SEP-1999; 99DE-01042095.
 PR 03-SEP-1999; 99DE-01042123.
 PR 03-SEP-1999; 99DE-01042125.
 XX
 XX (BADI) BASF AG.
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 XX
 DR WPI; 2001-061975/07.
 DR N-PSDB; AAF71706.
 XX
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes.
 XX
 XX Claim 20; Page 1129-1130; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243
 CC to AAB 79633 which are involved in carbon metabolism and energy
 CC production. The C. glutamicum SMP gene can be used in vectors (II) for
 CC expression in host cells and production or modulation of production of
 CC fine chemicals, such as, an organic acid, a proteinogenic or
 CC nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a
 CC nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid,
 CC a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a
 CC polypeptide, or an enzyme. The presence of (I) or SMP proteins (III)
 CC encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to C.
 CC glutamicum, identify and localise C. glutamicum sequences of interest, in
 CC evolutionary studies, in determining SMP protein regions required for
 CC function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH)
 XX
 SQ Sequence 285 AA;

Query Match 35.0%; Score 7; DB 4; Length 285;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFLL 11
 Db 141 SVLSFLL 147

```

RESULT 7
AAW29179
ID AAW29179 standard; protein; 355 AA.
XX
AC AAW29179;
XX
DT 19-DEC-1997 (first entry)
XX
DE Rat CC chemokine receptor.
XX
KW rat; CC chemokine receptor; screen; binding; ligand.
XX
OS Rattus rattus.
XX
PN JP09227599-A.
XX
PD 02-SEP-1997.
XX
PF 22-FEB-1996; 96JP-00035192.
XX
PR 22-FEB-1996; 96JP-00035192.
XX
PA (TAKE ) TAKEDA CHEM IND LTD.
XX
DR WPI; 1997-486426/45.
XX
DR N-PSDB; AAT86839.
XX
CC chemokine receptor protein - useful to screen for novel binding
PT compounds.
XX
PS Claim 1; Page 20-21; 26pp; Japanese.
XX
This sequence is a rat CC chemokine receptor. The receptor can be used to
CC screen for novel binding compounds and for preparation of antibodies or
CC antiserum
CC
SQ Sequence 355 AA;
Query Match 35.0%; Score 7; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLWTPY 15
Db 249 FLWTPY 255
|||||

RESULT 8
ADD45360
ID ADD45360 standard; protein; 355 AA.
XX
AC ADD45360;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rat Protein NP_065417, SEQ ID NO 10793.
XX
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
XX
PR 01-NOV-2001; 2001US-0346382P.
XX
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEHO ) GEN HOSPITAL CORP.

(FARB ) BAYER AG.
PA
XX
PI Woolf C, D'urso D, Befort K, Costigan M;
XX
DR WPI; 2003-268312/26.
XX
DR GENBANK; NP_065417.
XX
PT New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
XX
The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 355 AA;
Query Match 35.0%; Score 7; DB 7; Length 355;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLWTPY 15
Db 249 FLWTPY 255
|||||

RESULT 9
AAG92745
ID AAG92745 standard; protein; 1019 AA.
XX
AC AAG92745;
XX
DT 26-SRP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6499.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.

```


XX This invention describes novel human nucleic acid (cDNA) sequences (A),
 CC that are highly expressed in uterine tumour tissue and which have
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (ii) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (iii) for generation
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed
 CC sequence tags) from a particular tissue type before comparison of the
 CC expression patterns. This allows a significantly longer fragment of the
 CC gene to be revealed, so should reduce the number of failures associated
 CC with the fact that ESTs from different libraries may represent different
 CC parts of the same unknown gene, distorting the estimated frequency of
 CC occurrence in a particular tissue. AAY59941-Y60328 represent protein
 CC fragments encoded by the human endometrium tumour cDNA library derived
 CC EST fragments represented in AAZ41981-Z42121
 XX
 SQ Sequence 33 AA;

Query Match 30.0%; Score 6; DB 2; Length 33;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
 |||||
 Db 4 VLSFLL 9

RESULT 12

AAB64900
 ID AAB64900 standard; protein; 33 AA.

AC AAB64900;

XX 23-MAR-2001 (first entry)

XX Human secreted protein sequence encoded by gene 19 SEQ ID NO:78.

XX Human; secreted protein; diagnosis; immunomodulatory; anti-sclerotic;
 KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
 KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
 KW ophthalmological; neuroprotectant; neurotropic; anticonvulsant; vulnary;
 KW antialzheimers; antiparkinsonian; antimicrobial; immune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; HIV; infection;
 KW hyperproliferative disorder; cancer; Gaucher's disease; wound healing;
 KW cardiovascular disease; Scimitar syndrome; Chaga's cardiomyopathy;
 KW coronary arteriosclerosis; angiogenic disorder; diabetic retinopathy;
 KW corneal graft neovascularisation; neurological disorder; regeneration;
 KW Huntington's chorea; Alzheimer's disease; Parkinson's disease;
 KW infectious disease; chemotaxis.

XX Homo sapiens.

OS
 XX WO200076530-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US014933.

XX 11-JUN-1999; 99US-0138572P.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM, Komatsoulis GA;

XX WPI; 2001-071147/08.

DR N-PSDB; AAF33321.

XX Nucleic acids encoding 49 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's disease
 PT and diabetic retinopathy.

XX Claim 11; Page 494; 554pp; English.
 PS The polynucleotide sequences given in AAF33213 to AAF33261 encode the
 XX human secreted proteins given in AAB64882 to AAB64930. AAB64931 to
 CC AAB64991 represent human secreted polypeptide sequences and proteins
 CC homologous to them, which are given in the exemplification of the present
 CC invention. Human secreted proteins have activities based on the tissues
 CC and cells the genes are expressed in. Examples of activities include:
 CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
 CC antinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
 CC vascular; antimicrobial; anti-angiogenic; ophthalmological;
 CC neuroprotectant; anticonvulsant; neurotropic; antialzheimers;
 CC antiparkinsonian; and vulnary. The polynucleotides and polypeptides can
 CC be used in the prevention, diagnosis and treatment of diseases associated
 CC with inappropriate polypeptide expression. Disorders that may be
 CC prevented, diagnosed and/or treated by the above methods include immune
 CC disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
 CC human immuno-deficiency virus (HIV) infections), hyperproliferative
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
 CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
 CC neovascularisation and diabetic retinopathy), neurological disorders
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
 CC infectious diseases and/or for promoting wound healing, regeneration and
 CC /or chemotaxis. AAF33204 to AAF33212 and AAB64881 represent sequences
 CC used in the exemplification of the present invention
 XX
 SQ Sequence 33 AA;

Query Match 30.0%; Score 6; DB 4; Length 33;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
 |||||
 Db 3 VLSFLL 8

RESULT 13

AAB64424
 ID AAB64424 standard; protein; 33 AA.

XX AAB64424;

XX 23-MAR-2001 (first entry)

XX Human secreted protein sequence encoded by gene 3 SEQ ID NO:62.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
 KW angiogenesis; nervous system disorder; Alzheimer's disease; skin aging;
 KW ocular disorder; corneal infection; wound healing; food additive;
 KW preservative.

XX Homo sapiens.

XX WO200077255-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US014926.

XX 11-JUN-1999; 99US-0138628P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis GA;

XX

```

DR WPI; 2001-025337/03.
DR N-PSDB; AAF32701.
XX Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.
XX
PS Claim 11; Page 514; 593pp; English.
XX
CC The polynucleotide sequences given in AAF32699 to AAF32747 encode the
CC human secreted proteins given in AAF64422 to AAF64470. AAF64471 to
CC AAF64548 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC antiarthritic; immunosuppressive; antineumatic; antiproliferative;
CC cytoskeletal; cardiac; vasotropic; cerebroprotective; nootropic;
CC neuroprotective; antibacterial; virucide; fungicide; and
CC ophthalmological. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. AAF32699 to AAF32698 and
CC AAF64421 represent sequences used in the exemplification of the present
CC invention
XX
SQ Sequence 33 AA;
    Query Match          30.0%; Score 6; DB 4; Length 33;
    Best Local Similarity 100.0%; Pred. No. 41;
    Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    QY 6 VLGFLL 11
    Db 3 VLGFLL 8
        |||||
        |||||

RESULT 14
AAG24890
ID AAG24890 standard; protein; 38 AA.
XX
AC AAG24890;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 28737.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.

```

```

PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 04-JUN-1999; 99US-0137528P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 28-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.

```

PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145313P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.

PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0158294P.
PR 13-OCT-1999; 99US-0158295P.
PR 14-OCT-1999; 99US-0158329P.
PR 14-OCT-1999; 99US-0158330P.
PR 14-OCT-1999; 99US-0158331P.
PR 14-OCT-1999; 99US-0158637P.
PR 14-OCT-1999; 99US-0158638P.
PR 18-OCT-1999; 99US-0158584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 30.0%; Score 6; DB 3; Length 38;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
|||
Db 32 VLSFLL 37

RESULT 15

AAB44612

ID AAB44612 standard; protein; 41 AA.

XX AAB44612;

XX 12-FEB-2001 (first entry)

XX Human secreted protein sequence encoded by gene 17 SEQ ID NO:77.

Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
food additive; preservative; chromosome identification; cancer;
immune disorder; cardiovascular disorder; neurological disease;
wound healing; infectious disease.

XX Homo sapiens.

XX WO200058339-A2.

XX 05-OCT-2000.

XX 22-MAR-2000; 2000WO-US007440.

XX 26-MAR-1999; 99US-0126503P.

XX 17-DEC-1999; 99US-0172409P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX

DR WPI: 2000-594637/56.
DR N-PSDB; AAC79697.
XX
PT Fifty nucleic acid molecules encoding human secreted proteins, useful in
PT the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases.
XX
PS Claim 11; Page 375; 410pp; English.
XX
CC The polynucleotide sequences given in AAC79681 to AAC79730 encode the
CC human secreted proteins given in AAB44596 to AAB44645. AAB44646 to
CC AAB44693 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
CC antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
CC vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC cardiant. The polynucleotides and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides
CC can also be used as a food additive or preservative to increase or
CC decrease storage capabilities. The polynucleotides are useful for
CC chromosome identification. They are also useful as probes for diagnosing
CC a disorder related to the female reproductive system, particularly breast
CC and/or ovary cancer. They are also useful in the gene therapy of breast
CC and ovarian cancer. The nucleic acids, protein, antibodies, agonists and
CC antagonists from the present invention are useful in the diagnosis,
CC treatment and prevention of: cancer; immune disorders; cardiovascular
CC disorders; wound healing; neurological diseases; and infectious diseases.
CC AAC79672 to AAC79680 and AAB44595 represent sequences used in the
CC exemplification of the present invention
XX
SQ Sequence 41 AA;

Query Match 30.0%; Score 6; DB 3; Length 41;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LSFLW 12
|
|
|
|
|
|
|
Db 6 LSFLW 11

Search completed: July 19, 2004, 17:13:09
Job time : 57 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:12:07 ; Search time 18 Seconds
(without alignments)
57.362 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pbp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pbp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pbp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pbp.*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pbp.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pbp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	35.0	355	4	US-09-886-319A-13
2	6	30.0	33	4	US-09-673-395A-293
3	6	30.0	80	4	US-09-227-357-249
4	6	30.0	105	4	US-09-482-273-177
5	6	30.0	124	4	US-09-328-352-6057
6	6	30.0	148	4	US-09-489-039A-11733
7	6	30.0	154	4	US-09-313-458-34
8	6	30.0	154	4	US-09-746-359A-8
9	6	30.0	176	4	US-09-313-458-19
10	6	30.0	176	4	US-09-746-359A-5
11	6	30.0	215	4	US-09-252-991A-18224
12	6	30.0	338	4	US-09-107-532A-6781
13	6	30.0	345	4	US-09-482-273-113
14	6	30.0	369	4	US-09-482-273-208
15	6	30.0	377	1	US-08-153-848-42
16	6	30.0	377	3	US-09-299-843A-42
17	6	30.0	377	4	US-09-088-337B-42
18	6	30.0	377	5	PCT-US93-11153-42
19	6	30.0	408	4	US-09-107-532A-6563
20	6	30.0	426	4	US-09-134-000C-6607
21	6	30.0	431	4	US-09-543-681A-7623
22	6	30.0	739	4	US-09-134-001C-3586
23	5	25.0	8	1	US-08-419-594-3
24	5	25.0	8	2	US-08-919-089-3
25	5	25.0	9	4	US-09-311-784A-430
26	5	25.0	12	1	US-08-117-361C-12
27	5	25.0	13	4	US-09-522-666-30

```

28 5 25.0 15 4 US-09-009-953-49 Sequence 49, Appl
29 5 25.0 16 4 US-09-360-237-55 Sequence 55, Appl
30 5 25.0 16 4 US-09-680-571A-83 Sequence 83, Appl
31 5 25.0 16 4 US-09-680-571A-97 Sequence 97, Appl
32 5 25.0 17 1 US-08-487-890A-104 Sequence 104, App
33 5 25.0 17 2 US-08-478-435-104 Sequence 104, App
34 5 25.0 17 2 US-08-337-483-104 Sequence 104, App
35 5 25.0 17 2 US-08-478-373-104 Sequence 104, App
36 5 25.0 17 3 US-08-474-671-104 Sequence 104, App
37 5 25.0 17 3 US-08-483-577A-104 Sequence 104, App
38 5 25.0 17 3 US-08-897-438-104 Sequence 104, App
39 5 25.0 17 4 US-08-637-654-104 Sequence 104, App
40 5 25.0 17 4 US-08-649-518-104 Sequence 104, App
41 5 25.0 22 2 US-09-013-634-8 Sequence 104, App
42 5 25.0 27 4 US-09-360-237-57 Sequence 57, Appl
43 5 25.0 51 4 US-09-621-976-7274 Sequence 7274, Ap
44 5 25.0 53 4 US-09-663-600A-133 Sequence 133, App
45 5 25.0 54 4 US-09-621-976-5883 Sequence 5883, Ap

```

ALIGNMENTS

```

RESULT 1
US-09-886-319A-13
; Sequence 13, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886,319A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-886-319A-13

```

```

Query Match 35.0%; Score 7; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

```

```

QY 9 FLLWTPY 15
Db 249 FLLWTPY 255

```

```

RESULT 2
US-09-673-395A-293
; Sequence 293, Application US/09673395A
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE

```

```
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 293
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-673-395A-293

Query Match      30.0%; Score 6; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLL 11
DB      4 VLSFLL 9

RESULT 3
US-09-227-357-249
; Sequence 249, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953

; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 293
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-673-395A-293

Query Match      30.0%; Score 6; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLL 11
DB      4 VLSFLL 9

RESULT 4
US-09-482-273-177
; Sequence 177, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 177
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-482-273-177

Query Match      30.0%; Score 6; DB 4; Length 80;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLL 11
DB      15 VLSFLL 20

RESULT 4
US-09-482-273-177
; Sequence 177, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 177
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-482-273-177

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; LOCATION: (80)
; NAME/KEY: SITE
; FEATURE:
; LOCATION: (36)
; NAME/KEY: SITE
; FEATURE:
; ORGANISM: Homo sapiens
; TYPE: PRT
; LENGTH: 80
; SEQ ID NO 249
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
```

```

Query Match          30.0%; Score 6; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
   |||||
Db 26 VLSFLL 31

RESULT 5
US-09-328-352-6057
; Sequence 6057, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6057
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6057

Query Match          30.0%; Score 6; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
   |||||
Db 86 VLSFLL 91

RESULT 6
US-09-489-039A-11733
; Sequence 11733, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11733
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11733

Query Match          30.0%; Score 6; DB 4; Length 148;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
   |||||
Db 14 VLSFLL 19

RESULT 7
US-09-313-458-34
; Sequence 34, Application US/09313458
; Patent No. 6576743
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Haldeman, Betty A.

```

```

; TITLE OF INVENTION: MAMMALIAN CYTOKINE-LIKE POLYPEPTIDE-10
; FILE REFERENCE: 97-72
; CURRENT APPLICATION NUMBER: US/09/313,458
; CURRENT FILING DATE: 1999-05-17
; EARLIER APPLICATION NUMBER: 09/199,586
; EARLIER FILING DATE: 1998-11-25
; EARLIER APPLICATION NUMBER: 60/066,597
; EARLIER FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-313-458-34

Query Match          30.0%; Score 6; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FLLWTP 14
   |||||
Db 16 FLLWTP 21

RESULT 8
US-09-746-359A-8
; Sequence 8, Application US/09746359A
; Patent No. 6610286
; GENERAL INFORMATION:
; APPLICANT: Thompson, Penny
; APPLICANT: Foster, Donald C.
; APPLICANT: Xu, Wenfeng
; APPLICANT: Madden, Karen L.
; APPLICANT: Kelly, James D.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Blumberg, Hal
; APPLICANT: Eagan, Maribeth A.
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Chandrasekhar, Yasmin A.
; APPLICANT: No. 6610286ak, Julia E.
; TITLE OF INVENTION: Method for Treating Inflammation
; FILE REFERENCE: 99-108
; CURRENT APPLICATION NUMBER: US/09/746,359A
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/171,969
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/213,341
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-746-359A-8

Query Match          30.0%; Score 6; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FLLWTP 14
   |||||
Db 16 FLLWTP 21

RESULT 9
US-09-313-458-19
; Sequence 19, Application US/09313458
; Patent No. 6576743
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Haldeman, Betty A.

```

; TITLE OF INVENTION: MAMMALIAN CYTOKINE-LIKE POLYPEPTIDE-10
; FILE REFERENCE: 97-72
; CURRENT APPLICATION NUMBER: US/09/313,458
; CURRENT FILING DATE: 1999-05-17
; EARLIER APPLICATION NUMBER: 09/199,586
; EARLIER FILING DATE: 1998-11-25
; EARLIER APPLICATION NUMBER: 60/066,597
; EARLIER FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 19
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-313-458-19

Query Match 30.0%; Score 6; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 FLWTP 14
Db 16 FLWTP 21

RESULT 10
US-09-746-359A-5
; Sequence 5, Application US/09746359A
; Patent No. 6610286
; GENERAL INFORMATION:
; APPLICANT: Thompson, Penny
; APPLICANT: Foster, Donald C.
; APPLICANT: Xu, Wenfeng
; APPLICANT: Madden, Karen L.
; APPLICANT: Kelly, James D.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Blumberg, Hal
; APPLICANT: Eagan, Maribeth A.
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Chandrasekhar, Yasmin A.
; APPLICANT: No. 6610286ak, Julia E.
; TITLE OF INVENTION: Method for Treating Inflammation
; FILE REFERENCE: 99-108
; CURRENT APPLICATION NUMBER: US/09/746,359A
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/171,969
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/213,341
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-746-359A-5

Query Match 30.0%; Score 6; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 FLWTP 14
Db 16 FLWTP 21

RESULT 11
US-09-252-991A-18224
; Sequence 18224, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18224
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18224

Query Match 30.0%; Score 6; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 VLSFLL 11
Db 41 VLSFLL 46

RESULT 12
US-09-107-532A-6781
; Sequence 6781, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSER: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6781:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 338 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...338
; SEQUENCE DESCRIPTION: SEQ ID NO: 6781:

US-09-107-532A-6781

Query Match 30.0%; Score 6; DB 4; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
| | | | |
Db 302 VLSFLL 307

RESULT 13

US-09-482-273-113
; Sequence 113, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 113
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (53)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (345)
; OTHER INFORMATION: Xaa equals stop translation

US-09-482-273-113

Query Match 30.0%; Score 6; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
| | | | |
Db 265 VLSFLL 270

RESULT 14

US-09-482-273-208
; Sequence 208, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 208
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (78)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-482-273-208

Query Match 30.0%; Score 6; DB 4; Length 369;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
| | | | |
Db 290 VLSFLL 295

RESULT 15

US-08-153-848-42
; Sequence 42, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/153,848
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5759804and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-153-848-42

Query Match 30.0%; Score 6; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
| | | | |
Db 168 VLSFLL 173

Search completed: July 19, 2004, 17:15:15
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:14:36 ; Search time 42 Seconds
(without alignments)
148.838 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1285345 seqs, 312560633 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	95.0	20	14	US-10-195-730-116
2	7	35.0	57	11	US-09-864-408A-2008
3	7	35.0	93	12	US-10-424-599-165165
4	7	35.0	355	14	US-10-376-564-13
5	7	35.0	1019	9	US-09-738-626-6499
6	7	35.0	1019	15	US-10-380-055-12
7	35.0	1369	16	US-10-437-963-184093	
8	6	30.0	51	12	US-10-424-599-147607
9	6	30.0	54	12	US-10-424-599-148937
10	6	30.0	56	16	US-10-437-963-174252
11	6	30.0	62	11	US-09-864-408A-7962
12	6	30.0	62	12	US-10-424-599-187640
13	6	30.0	65	12	US-10-424-599-173594
14	6	30.0	68	16	US-10-437-963-146491
15	6	30.0	72	12	US-10-424-599-205684
					Sequence 116, App
					Sequence 2008
					Sequence 165165,
					Sequence 13, Appl
					Sequence 6499, App
					Sequence 12, Appl
					Sequence 184093,
					Sequence 17607,
					Sequence 148937,
					Sequence 174252,
					Sequence 7962, App
					Sequence 187640,
					Sequence 173594,
					Sequence 146491,
					Sequence 205684,

16	6	30.0	73	15	US-10-242-355-459	Sequence 459, App
17	6	30.0	77	12	US-10-424-599-281026	Sequence 281026,
18	6	30.0	79	12	US-09-973-278-260	Sequence 260, App
19	6	30.0	80	10	US-09-983-802-249	Sequence 249, App
20	6	30.0	80	12	US-09-984-490-249	Sequence 249, App
21	6	30.0	83	9	US-09-796-692-1812	Sequence 1812, Ap
22	6	30.0	83	9	US-09-796-692-2069	Sequence 2069, Ap
23	6	30.0	83	14	US-10-040-862-1812	Sequence 1812, Ap
24	6	30.0	83	14	US-10-040-862-2069	Sequence 2069, Ap
25	6	30.0	83	15	US-10-057-475B-1812	Sequence 1812, Ap
26	6	30.0	83	15	US-10-057-475B-2069	Sequence 2069, Ap
27	6	30.0	83	15	US-10-154-884B-1812	Sequence 1812, Ap
28	6	30.0	83	15	US-10-154-884B-2069	Sequence 2069, Ap
29	6	30.0	84	16	US-10-437-963-127297	Sequence 127297,
30	6	30.0	88	12	US-10-424-599-276079	Sequence 276079,
31	6	30.0	92	14	US-10-029-386-33935	Sequence 33935, A
32	6	30.0	99	10	US-09-764-891-3040	Sequence 3040, Ap
33	6	30.0	105	10	US-09-984-271-177	Sequence 177, App
34	6	30.0	105	12	US-09-984-276-177	Sequence 196176,
35	6	30.0	109	12	US-10-424-599-149648	Sequence 149648,
36	6	30.0	117	12	US-10-424-599-236063	Sequence 236063,
37	6	30.0	119	12	US-10-424-599-188984	Sequence 188984,
38	6	30.0	123	12	US-10-425-114-45129	Sequence 45129, A
39	6	30.0	141	12	US-10-424-599-188984	Sequence 188984,
40	6	30.0	149	12	US-10-425-114-50534	Sequence 50534, A
41	6	30.0	154	9	US-09-746-359A-8	Sequence 8, Appli
42	6	30.0	154	14	US-10-321-163-34	Sequence 34, Appl
43	6	30.0	154	14	US-10-413-661-34	Sequence 34, Appl
44	6	30.0	154	15	US-10-424-658-8	Sequence 8, Appli
45	6	30.0	154	16	US-10-471-151-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-10-195-730-116
; Sequence 116, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: PZ017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 116
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (20)
; OTHER INFORMATION: Xaa equals stop translation
US-10-195-730-116

Query Match 95.0%; Score 19; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYALKS 19

Db 1 MAHSVLSFLWTPYALKS 19

RESULT 2

```
US-09-864-408A-2008
; Sequence 2008, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2008
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-2008

Query Match          35.0%; Score 7; DB 11; Length 57;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFLL 11
Db 19 SVLSFLL 25

RESULT 3
US-10-424-599-165165
; Sequence 165165, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 165165
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_12015C.1.pap
US-10-424-599-165165

Query Match          35.0%; Score 7; DB 12; Length 93;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFLL 11
Db 35 SVLSFLL 41

RESULT 4
US-10-376-564-13
; Sequence 13, Application US/10376564
; Publication No. US20030180302A1
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckhard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound

US-09-864-408A-2008
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; FILE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014003
; CURRENT APPLICATION NUMBER: US/10/376,564
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 09/886,319
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 100 30 149.5-41
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-376-564-13

Query Match          35.0%; Score 7; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLWTPY 15
Db 249 FLWTPY 255

RESULT 5
US-09-738-626-6499
; Sequence 6499, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6499
; LENGTH: 1019
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6499

Query Match          35.0%; Score 7; DB 9; Length 1019;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFLL 11
Db 141 SVLSFLL 147

RESULT 6
US-10-380-055-12
; Sequence 12, Application US/10380055
```



```
; Publication No. US20040014180A1
; GENERAL INFORMATION:
; APPLICANT: BOTT, Michael, et al.
; TITLE OF INVENTION: A Method For Microbial Production Of Metabolic Products, Polynucle
; TITLE OF INVENTION: From Coryneform Bacteria And Their Use
; FILE REFERENCE: 032301 WN 331
; CURRENT APPLICATION NUMBER: US/10/380,055
; CURRENT FILING DATE: 2003-03-11
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1019
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-10-380-055-12

Query Match      35.0%; Score 7; DB 15; Length 1019;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 SVLSFLL 11
Db      141 SVLSFLL 147

RESULT 7
US-10-437-963-184093
; Sequence 184093, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 184093
; LENGTH: 1369
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_81119C.1.pep
; US-10-437-963-184093

Query Match      35.0%; Score 7; DB 16; Length 1369;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLLW 12
Db      742 VLSFLLW 748

RESULT 8
US-10-424-599-147607
; Sequence 147607, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
```

```
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 147607
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_104307C.1.pep
; US-10-424-599-147607

Query Match      30.0%; Score 6; DB 12; Length 51;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLL 11
Db      23 VLSFLL 28

RESULT 9
US-10-424-599-148937
; Sequence 148937, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 148937
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_105512C.1.pep
; US-10-424-599-148937

Query Match      30.0%; Score 6; DB 12; Length 54;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 SVLSFL 10
Db      18 SVLSFL 23

RESULT 10
US-10-437-963-174252
; Sequence 174252, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 174252
; LENGTH: 56
```

```
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_72210C.1.pep
US-10-437-963-174252

Query Match      30.0%; Score 6; DB 16; Length 56;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 LSFLW 12
        |||||
Db      47 LSFLW 52

RESULT 11
US-09-864-408A-7962
; Sequence 7962, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7962
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-7962

Query Match      30.0%; Score 6; DB 11; Length 62;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 SVLSFL 10
        |||||
Db      32 SVLSFL 37

RESULT 12
US-10-424-599-187640
; Sequence 187640, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 187640
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(62)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_140452C.1.pep
US-10-424-599-187640

Query Match      30.0%; Score 6; DB 12; Length 62;
```

```
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLL 11
        |||||
Db      32 VLSFLL 37

RESULT 13
US-10-424-599-173594
; Sequence 173594, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 173594
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_127775C.1.pep
US-10-424-599-173594

Query Match      30.0%; Score 6; DB 12; Length 65;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 LSFLW 12
        |||||
Db      46 LSFLW 51

RESULT 14
US-10-437-963-146491
; Sequence 146491, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 146491
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_4710C.1.pep
US-10-437-963-146491

Query Match      30.0%; Score 6; DB 16; Length 68;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 LSFLW 12
        |||||
```

Db 50 LSFLW 55

RESULT 15
US-10-424-599-205684
; Sequence 205684, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 205684
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_27760C.1.pap
US-10-424-599-205684

Query Match 30.0%; Score 6; DB 12; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 AHSVLS 8
| | | | |
Db 54 AHSVLS 59

Search completed: July 19, 2004, 17:19:49
Job time : 42 secs

This Page Blank (uspto)

Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 HSVLSFL 10
Db 178 HSVLSFL 184

RESULT 3
W9BPC7
gene 19.3 protein - phase T7
C:Species: phase T7
C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 23-Jul-1999
C:Accession: A04415; S42340
R:Dunn, J.J.; Thompson, K.
submitted to the Nucleic Acid Sequence Database, September 1982
A:Reference number: A94615
A:Accession: A04415
A:Molecule type: DNA
A:Residues: 1-57 <DUN>
R:Dunn, J.J.; Studier, F.W.
J. Mol. Biol. 186, 477-535, 1983
A:Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 gene
A:Reference number: S42283; MUID:83241725; PMID:6864790
A:Accession: S42340
A:Molecule type: DNA
A:Residues: 1-57 <DUN>
A:Cross-references: EMBL:V01146; NID:g431187; PIDN:CAA24442.1; PID:g15618
A:Note: the authors did not translate the codon for residue 1
C:Genetics:
A:Gene: 19.3
A:Map position: 96.53-96.96
C:Superfamily: phase T7 gene 19.3 protein

Query Match 30.0%; Score 6; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
Db 38 SVLSFL 43

RESULT 4
T07432
Photosystem II protein psbX - Japanese black pine chloroplast
C:Species: chloroplast Pinus thunbergiana (Japanese black pine)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
C:Accession: T07432
R:Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugiyura, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A:Title: Loss of all ndh genes as determined by sequencing the entire chloroplast genome
A:Reference number: Z16030; MUID:95024047; PMID:7937893
A:Accession: T07432
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-59 <WAK>
A:Cross-references: EMBL:D17510; NID:g529643; PIDN:BAA04312.1; PID:g1262593
C:Genetics:
A:Gene: psbX
C:Superfamily: photosystem II protein psbX
C:Keywords: chloroplast; membrane-associated complex; photosynthesis; photosystem II; th

Query Match 30.0%; Score 6; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
Db 43 VLSFLL 48

RESULT 5
T06236
probable 3,4-dihydroxy-2-butanone kinase - soybean (fragment)
C:Species: Glycine max (soybean)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 08-Sep-2000
C:Accession: T06236
R:Mahalingam, R.; Knap, H.T.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z15555
A:Accession: T06236
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-82 <MAH>
A:Cross-references: EMBL:AF047054; NID:g2905783; PIDN:AAC03560.1; PID:g2905784
A:Experimental source: cultivar Essex; root
C:Genetics:
A:Gene: dbk
C:Superfamily: glycerone kinase DAK1

Query Match 30.0%; Score 6; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAHSV 7
Db 32 AAHSV 37

RESULT 6
F87485
flhB-related protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: F87485
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hart, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87485
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-87 <STO>
A:Cross-references: GB:AE005673; NID:g13423358; PIDN:AAK23882.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1907

Query Match 30.0%; Score 6; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
Db 80 VLSFLL 85

RESULT 7
C72852
Acorf-19 protein - Autographa californica nuclear polyhedrosis virus
C:Species: Autographa californica nuclear polyhedrosis virus, ACMPV
A:Note: daDNA virus
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-Jan-2000
C:Accession: C72852
R:Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.
Virology 202, 586-605, 1994
A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus
A:Reference number: A72850; MUID:94303173; PMID:8030224
A:Accession: C72852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <AYR>
A:Cross-references: GB:L22858; NID:g510708; PIDN:AAA66649.1; PID:g559088

C;Genetics:
A;Gene: ACorf-19
C;Superfamily: AcMNPV hypothetical protein 19

Query Match 30.0%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
|
|
|
|
|
Db 99 VLSFLL 104

RESULT 8
D44221
orf4 protein - Autographa californica nuclear polyhedrosis virus
C;Species: Autographa californica nuclear polyhedrosis virus, AcMNPV
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 11-Jan-2000
C;Accession: D44221
R;Braunagel, S.C.; Daniel, K.D.; Reilly, L.M.; Guarino, L.A.; Hong, T.; Summers, M.D.
Virology 191, 1003-1008, 1992
A;Title: Sequence, genomic organization of the EcoRI-A fragment of Autographa californica VP8 of rotavirus.
A;Reference number: A44221; MUID:93079853; PMID:1333113
A;Accession: D44221
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-108 <BRA>
A;Cross-references: GB:S52569
C;Superfamily: AcMNPV hypothetical protein 19

Query Match 30.0%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
|
|
|
|
|
Db 99 VLSFLL 104

RESULT 9
T41765
AcMNPV orf19 - Bombyx mori nuclear polyhedrosis virus (isolate T3)
C;Species: Bombyx mori nuclear polyhedrosis virus, BmSNPV
A;Variety: isolate T3
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: T41765
R;Gomi, S.; Majima, K.; Maeda, S.
J. Gen. Virol. 80, 1323-1337, 1999
A;Title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.
A;Reference number: 222020; MUID:99281911; PMID:10355780
A;Accession: T41765
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-110 <KAM>
A;Cross-references: EMBL:L33180; NID:g3745835; PIDN:AAC63694.1; PID:g3745847
A;Experimental source: isolate T3
C;Genetics:
A;Note: Orf 11
C;Superfamily: AcMNPV hypothetical protein 19

Query Match 30.0%; Score 6; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
|
|
|
|
|
Db 101 VLSFLL 106

RESULT 10
F71032
hypothetical protein PH1550 - Pyrococcus horikoshii

C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: F71032
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekinaka, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, D.N.S. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon.
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71032
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-111 <KAW>
A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30662.1; PID:g3257979
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1550
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1550

Query Match 30.0%; Score 6; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LSFLW 12
|
|
|
|
|
Db 103 LSFLW 108

RESULT 11
H90450
hypothetical protein SSO2753 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: H90450
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayer, M.J.; Chan-Ying, J.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.; arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: H90450
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-180 <KUR>
A;Cross-references: GB:AE006641; NID:gl3816084; PIDN:AAK42863.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO2753

Query Match 30.0%; Score 6; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
|
|
|
|
|
Db 73 VLSFLL 78

RESULT 12
B83183
hypothetical protein PA3698 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83183
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Briadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen.
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83183
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-183 <STO>

A;Cross-references: GB:AE004789; GB:AE004091; NID:g9949857; PIDN:AAG07086.1; GSPDB:GN001
 A;Experimental source: strain PA01
 C;Genetics:
 A;Gene: FA3698

Query Match 30.0%; Score 6; DB 2; Length 183;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFL 11
 |||||
 Db 9 VLSFL 14

RESULT 13
 F85515
 unknown protein encoded in prophage CP-933I [imported] - Escherichia coli (strain O157:H7)
 C;Species: Escherichia coli
 C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C;Accession: F85515

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamocousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551
 A;Accession: F85515
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-205 <STO>
 A;Cross-references: GB:AE005174; NID:gil2513038; PIDN:AAG54586.1; GSPDB:GN00145; UWGP:Z03
 A;Experimental source: strain O157:H7, substrain EDL933
 C;Genetics:
 A;Gene: Z0326

Query Match 30.0%; Score 6; DB 2; Length 205;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AHSVLS 8
 |||||
 Db 174 AHSVLS 179

RESULT 14
 C90665
 hypothetical protein ECs0291 [imported] - Escherichia coli (strain O157:H7, substrain R)
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: C90665

R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: C90665
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-205 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BAE33714.1; PID:gil3359748; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RMD 0509952
 C;Genetics:
 A;Gene: ECs0291

Query Match 30.0%; Score 6; DB 2; Length 205;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AHSVLS 8
 |||||
 Db 174 AHSVLS 179

RESULT 15

AH3615
 gdp-4-dehydro-D-rhamnose reductase (EC 1.1.1.187) [imported] - Brucella melitensis (stra
 C;Species: Brucella melitensis
 C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C;Accession: AH3615
 R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
 ; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagijs, S.; O'Callaghan, D.; Letess
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
 A;Reference number: AD3252; PMID:11756688
 A;Accession: AH3615
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-246 <KUR>
 A;Cross-references: GB:AE008918; PIDN:AAL54091.1; PID:gi17985049; GSPDB:GN00191
 A;Experimental source: strain 16M

C;Genetics:
 A;Gene: EMEI10849
 A;Map position: II
 C;Keywords: oxidoreductase

Query Match 30.0%; Score 6; DB 2; Length 246;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAHSV 7
 |||||
 Db 47 AAHSV 52

Search completed: July 19, 2004, 17:14:46
 Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:07:32 ; Search time 13 Seconds
(without alignments)
80.108 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	35.0	355	1 CKR1_MOUSE	P51675 mus musculus
2	6	30.0	56	1 PSBK_PINTH	P41598 pinus thunb
3	6	30.0	57	1 V193_BPT7	P03790 bacterioph
4	6	30.0	82	1 DHBK_SOYEN	O49227 glycine max
5	6	30.0	92	1 YB41_STRMU	P59154 streptococ
6	6	30.0	108	1 Y019_NPVAC	P41424 autographa
7	6	30.0	156	1 ARQ1_RALSO	Q8XVP5 ralstonia s
8	6	30.0	176	1 IL20_MOUSE	Q9JKV9 mus musculu
9	6	30.0	285	1 PSD_VIBPA	Q87K29 vibrio para
10	6	30.0	286	1 RIPI_MOMCH	P16094 momordica c
11	6	30.0	297	1 CTA4_BACPF	Q04443 bacillus ps
12	6	30.0	317	1 O6N2_HUMAN	Q8NGY6 homo sapien
13	6	30.0	319	1 O6K6_HUMAN	Q8NGW6 homo sapien
14	6	30.0	322	1 O5T3_HUMAN	Q8NGG3 homo sapien
15	6	30.0	330	1 O808_MOUSE	Q8VFI3 mus musculu
16	6	30.0	344	1 LPXK_NEIMA	Q9JVE4 neisseria m
17	6	30.0	344	1 LPXK_NEIME	Q9K0K7 neisseria m
18	6	30.0	342	1 ZDH4_HUMAN	Q9NP98 homo sapien
19	6	30.0	352	1 L4R1_HUMAN	Q15722 homo sapien
20	6	30.0	377	1 OPS2_HEMSA	Q25158 hemigrapsus
21	6	30.0	456	1 ENTI_RAT	O54698 rattus norv
22	6	30.0	458	1 TCR_BACST	P07561 bacillus st
23	6	30.0	458	1 TCR_STRAG	P36890 streptococ
24	6	30.0	458	1 TCR_STRAG	P13924 streptococ
25	6	30.0	458	1 TCR_STRPN	P11063 streptococ
26	6	30.0	461	1 PEX2_PICPA	Q01964 picchia past
27	6	30.0	477	1 DTAL_RHILO	Q98AV2 rhizobium l
28	6	30.0	483	1 CTDB_HUMAN	Q9H579 homo sapien
29	6	30.0	492	1 YCLF_BACSU	P94408 bacillus su
30	6	30.0	494	1 KCF1_HUMAN	Q9H3M0 homo sapien
31	6	30.0	513	1 AMY3_SCHPO	O14154 schizosacch
32	6	30.0	552	1 YGXB_BACSU	P37874 bacillus su
33	6	30.0	565	1 MOT8_MOUSE	O70324 mus musculu

```

34 6 30.0 579 1 FD21_MYCLE P54200 mycobacteri
35 6 30.0 613 1 MOT8_HUMAN P36021 homo sapien
36 6 30.0 649 1 QOX1_BACSU P34956 bacillus su
37 6 30.0 790 1 LY14_YEAST P40971 saccharomyc
38 6 30.0 859 1 CHS1_SCHPO P30597 schizosacch
39 6 30.0 1324 1 SMC4_SCHPO P41004 schizosacch
40 6 30.0 1416 1 BLM_MOUSE O88700 mus musculu
41 6 30.0 1556 1 PRO5_DROVI Q9U6A1 drosophila
42 6 30.0 1620 1 CO3_EPTBU P98094 eptaretus
43 6 30.0 2103 1 RRPL_UUK P33453 ukuniemi v
44 6 30.0 3859 1 TRAP_HUMAN Q9Y4A5 homo sapien
45 5 25.0 56 1 NU6M_ARTSA P19048 artemia sal

```

ALIGNMENTS

```

RESULT 1
CKR1_MOUSE
ID CKR1_MOUSE STANDARD; PRT; 355 AA.
AC P51675; Q91VP9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE C-C chemokine receptor type 1 (C-C CKR-1) (CCR-1) (CCR1)
DE (Macrophage inflammatory protein-1 alpha receptor) (MIP-1alpha-R)
DE (RANTES-R).
DE CCR1 OR CMKBR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv; TISSUE=Peritoneal macrophage;
RX MEDLINE=96072806; PubMed=7594543;
RA Post T.W., Bozic C.R., Rothenberg M.E., Luster A.D., Gerard N.,
RA Gerard C.;
RT "Molecular characterization of two murine eosinophil beta chemokine
RT receptors.";
RL J. Immunol. 155:5299-5305(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX MEDLINE=95340546; PubMed=7542241;
RA Gao J.-L., Murphy P.M.;
RT "Cloning and differential tissue-specific expression of three mouse
RT beta chemokine receptor-like genes, including the gene for a
RT functional macrophage inflammatory protein-1 alpha receptor.";
RL J. Biol. Chem. 270:17494-17501(1995).
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RX MEDLINE=22398257; PubMed=12477932;
RA Strausberg R.L., Reingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Tashiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McWray P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shervencko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";

```

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: RECEPTOR FOR A C-C TYPE CHEMOKINE. BINDS TO MIP-1-ALPHA, RANTES, AND LESS EFFICIENTLY, TO MIP-1-BETA OR MCP-1 AND SUBSEQUENTLY TRANSDUCES A SIGNAL BY INCREASING THE INTRACELLULAR CALCIUM IONS LEVEL. RESPONSIBLE FOR AFFECTING STEM CELL PROLIFERATION.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- TISSUE SPECIFICITY: Detected in the heart, spleen, lung, peritoneal exudate cells and leukocytes.

CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U29678; AAA86119.1; -

DR EMBL; U28404; AAA89153.1; -

DR EMBL; BC011092; AAH11092.1; -

DR F1R; I49339; I49339.

DR MGD; MGI:104618; Ccrl.

DR GO; GO:0016493; F:C-C chemokine receptor activity; IDA.

DR GO; GO:0005515; F:protein binding; IPI.

DR GO; GO:0030595; P:immune cell chemotaxis; IDA.

DR GO; GO:0006954; P:inflammatory response; IMP.

DR GO; GO:0030099; P:myeloid blood cell differentiation; IMP.

DR InterPro; IPR000276; GPCR_Rhodapn.

DR Pfam; PF00001; 7tm 1; 1.

DR PRINTS; PR00237; GPCR_RHODOPSIN.

DR PROSITE; PS00237; G PROTEIN RECP F1_1; 1.

DR PROSITE; PS0262; G PROTEIN RECP F1_2; 1.

KW G-protein coupled receptor; Transmembrane.

FT DOMAIN 1 34 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 35 60 1 (POTENTIAL).

FT DOMAIN 61 64 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 65 91 2 (POTENTIAL).

FT DOMAIN 92 107 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 108 129 3 (POTENTIAL).

FT DOMAIN 130 146 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 147 171 4 (POTENTIAL).

FT DOMAIN 172 197 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 198 223 5 (POTENTIAL).

FT DOMAIN 224 239 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 240 264 6 (POTENTIAL).

FT DOMAIN 265 281 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 282 305 7 (POTENTIAL).

FT DOMAIN 306 355 CYTOPLASMIC (POTENTIAL).

FT DISULFID 106 183 BY SIMILARITY.

FT CONFLICT 55 55 M -> V (IN REF. 2 AND 3).

FT CONFLICT 149 149 L -> F (IN REF. 3).

FT CONFLICT 278 278 H -> Q (IN REF. 3).

SQ SEQUENCE 355 AA; 40901 MW; FCE9FF70E6F38B1 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FILWTPY 15
Db 249 FILWTPY 255
|||||

RESULT 2

ID PSBK_PINTH STANDARD; PRT; 56 AA.

AC P41598;

DT 01-NOV-1995 (Rel. 32, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Photosystem II reaction center protein K precursor (PSII-K).

GN PSBK.

OS Pinus thunbergii (Green pine) (Japanese black pine).

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.

OX NCBI_TaxID=3350;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95024047; PubMed=7937893;

RA Waksugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T., Sugitara M.;

RT "Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine Pinus thunbergii.";

RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).

CC -!- FUNCTION: This protein is a component of the reaction center of photosystem II.

CC -!- SIMILARITY: Belongs to the psbK family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; DJ7510; BAA04312.1; ALT_INIT.

DR HAMAP; MF_00441; -; 1. PSII_PsbK.

DR InterPro; IPR003687; PSII_PsbK.

DR Pfam; PF02533; PsbK; 1.

KW Photosystem II; Chloroplast.

FT PROPEP 1 19 POTENTIAL.

FT CHAIN 20 56 PHOTOSYSTEM II REACTION CENTER PROTEIN K.

SQ SEQUENCE 56 AA; 6346 MW; 18B1DEEF198ACASA CRC64;

Query Match 30.0%; Score 6; DB 1; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSELL 11
Db 40 VLSELL 45
|||||

RESULT 3

ID V193_BPT7 STANDARD; PRT; 57 AA.

AC P03790;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 01-MAR-1989 (Rel. 10, Last annotation update)

DE Gene 19.3 protein.

GN 19.3.

OS Bacteriophage T7.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;

OC T7-like viruses.

OX NCBI_TaxID=10760;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=83241725; PubMed=6864790;

RA Dunn J.J., Studier F.W.;

RT "Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements.";

RL J. Mol. Biol. 166:477-535(1983).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

```
DR EMBL; V01146; CAA24442.1; -.
DR PIR; A04415; W9BPC7.
SQ SEQUENCE 57 AA; 5561 MW; A16EA477F5203860 CRC64;

Query Match      30.0%; Score 6; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
Db 38 SVLSFL 43

RESULT 4
ID DBK_SOYEN STANDARD; PRT; 82 AA.
AC O49227;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative 3,4-dihydroxy-2-butanone kinase (EC 2.7.1.-) (Fragment).
GN DBK.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Essex; TISSUE=Root;
RA Herz S., Eberhardt S., Bacher A.;
RT "L. esculentum mRNA for 3,4-dihydroxy-2-butanone kinase.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the dihydroxyacetone kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF047054; AAC03560.1; -.
DR PIR; T06236; T06236.
DR InterPro; IPR004007; Dak2.
DR Pfam; PF02734; Dak2; 1.
KW Transferase; Kinase.
FT NON TER 1 1
FT NON TER 82 82
SQ SEQUENCE 82 AA; 8177 MW; 69BCC70B61210BA0 CRC64;

Query Match      30.0%; Score 6; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAHSVIL 7
Db 32 AAHSVIL 37

RESULT 5
ID YB41 STRMU STANDARD; PRT; 92 AA.
AC P59154;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein SMU.1141c.
GN SMU.1141c.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.

OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- SIMILARITY: Belongs to the UPF0223 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE014951; AAN5833.1; -.
DR HAMAP; MF 01041; -.
DR InterPro; IPR007920; UPF0223.
DR Pfam; PF05256; UPF0223; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 92 AA; 10622 MW; BBC44A96FEF0BD1 CRC64;

Query Match      30.0%; Score 6; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
Db 20 SVLSFL 25

RESULT 6
ID Y019 NPVAC STANDARD; PRT; 108 AA.
AC P41424;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 12.2 kDa protein in EGT-IAP1 intergenic region (ORF4).
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OX Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OX Nucleopolyhedrovirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL Virology 202:586-605(1994).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=E2;
CC MEDLINE=93079853; PubMed=1333113;
CC Braunagel S.C., Daniel K.D., Reilly L.M., Guarino L.A., Hong T.,
CC Summers M.D.;
CC "Sequence, genomic organization of the EcoRI-A fragment of Autographa
CC californica nuclear polyhedrosis virus, and identification of a
CC viral-encoded protein resembling the outer capsid protein VP8 of
CC rotavirus.";
CC Virology 191:1003-1008(1992).
CC -!- SIMILARITY: TO CORRESPONDING ORF IN OPMNPV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)

CC -----
CC EMBL; L22858; AAA66649.1; -;
CC DR EMBL; M96361; AAA66789.1; -;
CC DR PIR; C72852; C72852.
CC DR PIR; D44221; D44221.
CC KW Hypothetical protein.
CC FT CONFLICT 6
CC SQ SEQUENCE 108 AA; 12162 MW; 8363B15DF72486A1 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
Db 99 VLSFLL 104
|||||

RESULT 7

ARQ1 RALSO
ID ARQ1 RALSO STANDARD; PRT; 156 AA.
AC Q8XVP5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 3-dehydroquininate dehydratase 1 (EC 4.2.1.10) (3-dehydroquinase 1)
(Type II DHQase 1).
GN ARQ01 OR RSC2785 OR R500062.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salancoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).

CC -1- FUNCTION: Catalyzes a trans-dehydration via an enolate
intermediate (By similarity).
CC -1- CATALYTIC ACTIVITY: 3-dehydroquininate = 3-dehydroshikimate + H(2)O.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
third step.
CC -1- SIMILARITY: Belongs to the type-II 3-dehydroquinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)

CC -----
CC EMBL; AL646071; CAD16492.1; -;
CC DR HAMAP; MF_00169; -; 1
CC DR InterPro; IPR001874; DHQuinase_II.
CC DR Pfam; PF01220; DHQuinase_II; 1.
CC DR ProDom; PD004527; DHQuinase_II; 1.
CC DR TIGRfams; TIGR01088; aroQ; 1.
CC DR PROSITE; PS01029; DEHYDROQUINASE_II; 1.
CC KW Aromatic amino acid biosynthesis; Lyase; Complete proteome.
CC SQ SEQUENCE 156 AA; 16521 MW; 86BC38F65A94D032 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAHSV 7
Db 9 AAHSV 14
|||||

RESULT 8

IL20_MOUSE
ID IL20_MOUSE STANDARD; PRT; 176 AA.
AC Q9JKV9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Interleukin-20 precursor (IL-20) (Four alpha helix cytokine ZCYT010).
GN IL20.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=21097717; PubMed=11163236;
RA Blumberg H., Conklin D., Xu W.F., Grossmann A., Brender T.,
RA Carollo S., Eagan M., Foster D., Haldeman B.A., Hammond A., Haugen H.,
RA Jelinek L., Kelly J.D., Madden K., Maurer M.F., Parrish-Novak J.,
RA Prunkard D., Saxon S., Spracher C., Waggle K., West J.,
RA Whitmore T.E., Yao L., Kuechle M.K., Dale B.A., Chandrasekhar Y.A.;
RT "Interleukin 20: discovery, receptor identification, and role in
epidermal function";
RL Cell 104:9-19(2001).

CC -1- FUNCTION: Cytokine that may be involved in epidermal function and
psoriasis. Acts through STAT3.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the IL-10 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)

CC -----
CC EMBL; AF224267; AAF36680.1; -;
CC DR MGD; MGI:1890473; IL20.
CC DR InterPro; IPR000098; Interleukin_10.
CC DR ProDom; PD003687; Interleukin_10; 1.
CC DR SMART; SM00188; IL10; 1.
CC DR PROSITE; PS00520; INTERLEUKIN_10; 1.
CC KW Cytokine; Signal.
CC FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 176 INTERLEUKIN-20.
FT DISULFID 33 126 POTENTIAL.
FT DISULFID 80 132 POTENTIAL.
FT DISULFID 81 134 POTENTIAL.
SQ SEQUENCE 176 AA; 20097 MW; 08577AF656574771 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLLWTP 14
Db 16 FLLWTP 21
|||||

RESULT 9

PSD_VIBPA
ID PSD_VIBPA STANDARD; PRT; 285 AA.

AC Q87KZ9;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phosphatidylserine decarboxylase proenzyme (EC 4.1.1.65) [Contains:
 DE Phosphatidylserine decarboxylase alpha chain; Phosphatidylserine
 DE decarboxylase beta chain].
 GN PSD OR VP2825.
 OS Vibrio parahaemolyticus.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIND 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distinct from that of V. cholerae.";
 RL Lancet 361:743-749(2003).
 CC -!- CATALYTIC ACTIVITY: Phosphatidyl-L-serine =
 CC phosphatidylethanolamine + CO(2).
 CC -!- COFACTOR: Pyruvoyl group (By similarity).
 CC -!- SIMILARITY: Belongs to the phosphatidylserine decarboxylase
 CC family. Subfamily 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AP005082; BAC61088.1; -;
 DR HAMAP; MF 00662; -; 1.
 DR InterPro; IPR003817; PS_Dcarboxylase.
 DR Pfam; PF02666; PS_Dcarboxylase; 1.
 KW Phospholipid biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen;
 KW Complete proteome.
 FT CHAIN 1 251 PHOSPHATIDYL SERINE DECARBOXYLASE BETA
 FT CHAIN (BY SIMILARITY).
 FT CHAIN 252 285 PHOSPHATIDYL SERINE DECARBOXYLASE ALPHA
 FT CHAIN (BY SIMILARITY).
 FT SITE 251 252 CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY).
 FT MOD_RES 252 252 CONVERTED TO A PYRUVOYL GROUP (BY
 FT SIMILARITY).
 FT SEQUENCE 285 AA; 31730 MW; 9CD9C934BFAB4EE2 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 285;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 13 TPVALK 18
 |||||
 Db 279 TPVALK 284
 RESULT 10
 RIPI_MOMCH STANDARD; PRT; 286 AA.
 AC P16094; P24697;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ribosome-inactivating protein monomeric precursor (EC 3.2.2.22) (rRNA
 DE N-glycosidase) (Alpha-momorcharin) (Alpha-MMC).
 OS Momordica charantia (Bitter melon) (Balsam pear).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 CC eurosids I; Cucurbitales; Cucurbitaceae; Momordica.

OX NCBI_TaxID=3673;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seed;
 RX MEDLINE=91159486; PubMed=2001404;
 RA Ho W.K.K., Liu S.C., Shaw P.C., Yeung H.W., Ng T.B., Chan W.Y.;
 RT "Cloning of the cDNA of alpha-momorcharin: a ribosome inactivating
 RT protein.";
 RL Biochim. Biophys. Acta 1088:311-314(1991).
 RN [2]
 RP SEQUENCE OF 24-38.
 RC TISSUE=Seed;
 RX MEDLINE=89326691; PubMed=2753596;
 RA Montecucchi P.C., Lazzarini A.M., Barbieri L., Stirpe F., Soria M.,
 RA Lippi D.;
 RT "N-terminal sequence of some ribosome-inactivating proteins.";
 RL Int. J. Pept. Protein Res. 33:263-267(1989).
 RN [3]
 RP SEQUENCE OF 24-70.
 RC TISSUE=Seed;
 RX MEDLINE=89005108; PubMed=3262509;
 RA Casellas P., Dussosoy D., Falasca A.I., Barbieri L.,
 RA Guillemot J.C., Ferrara P., Bolognesi A., Cenini P., Stirpe F.;
 RT Trichokirin, a ribosome-inactivating protein from the seeds of
 RT Trichosanthes kirilowii Maximowicz. Purification, partial
 RT characterization and use for preparation of immunotoxins.";
 RL Eur. J. Biochem. 176:581-588(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=94356447; PubMed=8075985;
 RA Ren J., Wang Y., Dong Y., Stuart D.I.;
 RT "The N-glycosidase mechanism of ribosome-inactivating proteins
 RT implied by crystal structures of alpha-momorcharin.";
 RL Structure 2:7-16(1994).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.16 ANGSTROMS).
 RX MEDLINE=94192822; PubMed=8143869;
 RA Husain J., Tickle I.J., Wood S.P.;
 RT "Crystal structure of momordin, a type I ribosome inactivating
 RT protein from the seeds of Momordica charantia.";
 RL FEBS Lett. 342:154-158(1994).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=95344383; PubMed=7619070;
 RA Huang Q., Liu S., Tang Y., Jin S., Wang Y.;
 RT "Studies on crystal structures, active-centre geometry and
 RT depurinating mechanism of two ribosome-inactivating proteins.";
 RL Biochem. J. 309:285-298(1995).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of the N-glycosidic bond at one
 CC specific adenosine on the 28S rRNA.
 CC -!- SIMILARITY: Belongs to the ribosome-inactivating protein family.
 CC Type 1 RIP subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X57682; CAA0869.1; -;
 DR PIR; S14273; RLPUGG.
 DR PDB; 1AHA; 22-JUN-94.
 DR PDB; 1AHB; 22-JUN-94.
 DR PDB; 1AHC; 22-JUN-94.
 DR PDB; 1MOM; 31-MAY-94.
 DR PDB; 1MRG; 07-FEB-95.
 DR PDB; 1MRH; 07-FEB-95.
 DR PDB; 1MRI; 07-FEB-95.
 DR PDB; 1F8Q; 21-NOV-01.
 DR GlycoSuiteDB; P16094; -;
 DR InterPro; IPR001574; RIP.

DR Pfam; PF00161; RIP; 1.
DR PRINTS; PR00396; SHIGARICIN.
DR PROSITE; PS00275; SHIGA_RICIN; 1.
KW Plant defense; Protein synthesis inhibitor; Hydrolase; Toxin; Signal;
FT Glycoprotein; 3D-structure.
FT SIGNAL 1 23
FT CHAIN 24 269 RIBOSOME-INACTIVATING PROTEIN MOMORDIN I.
FT PROPP 270 286 MISSING IN MATURE PROTEIN.
FT ACT_SITE 183 183
FT CARBOHYD 250 250 N-LINKED (GLNAC. . .).
FT FTID=CAR_000082.
FT STRAND 25 28
FT HELIX 34 47
FT STRAND 50 54
FT TURN 55 56
FT STRAND 57 60
FT HELIX 66 69
FT STRAND 70 76
FT TURN 78 79
FT STRAND 82 88
FT TURN 89 92
FT STRAND 93 99
FT TURN 100 101
FT STRAND 102 105
FT HELIX 109 114
FT TURN 115 117
FT TURN 120 121
FT STRAND 124 127
FT HELIX 134 141
FT HELIX 145 147
FT STRAND 150 150
FT HELIX 152 162
FT TURN 163 163
FT HELIX 167 186
FT STRAND 187 187
FT HELIX 188 195
FT TURN 196 197
FT STRAND 202 202
FT HELIX 206 225
FT TURN 226 230
FT STRAND 231 238
FT TURN 240 241
FT STRAND 246 250
FT TURN 251 252
FT HELIX 254 258
FT TURN 259 259
FT STRAND 260 260
FT STRAND 263 263
FT HELIX 266 268
SQ SEQUENCE 286 AA; 31532 MW; E1B013ABEBC216CF CRC64;
Query Match 30.0%; Score 6; DB 1; Length 286;
Best Local Similarity 100.0%; Pred.No.24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 SVLSFL 10
Db 5 SVLSFL 10
RESULT 11
CTAA_BACPF STANDARD; PRT; 297 AA.
AC Q0443;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cytochrome Aa3 controlling protein.
GN CTAA.
OS Bacillus pseudofirmus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=79885;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=OF4;
RX MEDLINE=93107080; PubMed=7678007;
RA Quirk P.G., Hicks D.B., Krulwich T.A.;
RT "Cloning of the cta operon from alkaliphilic Bacillus firmus OF4 and
RT characterization of the pH-regulated cytochrome caa3 oxidase it
RT encodes."
RL J. Biol. Chem. 268:678-685(1993).
CC -!- FUNCTION: CTAA IS REQUIRED FOR CYTOCHROME Aa3 BIOSYNTHESIS AND
CC INDEPENDENTLY FOR SPOULATION.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M94110; AAA22362.1; -.
DR PIR; A45335; A45335.
DR InterPro; IPR003780; COX15_Ctaa.
DR Pfam; PF02628; COX15-Ctaa; 1.
KW Sporulation; Membrane.
SQ SEQUENCE 297 AA; 32326 MW; 1488B374029C0E59 CRC64;
Query Match 30.0%; Score 6; DB 1; Length 297;
Best Local Similarity 100.0%; Pred.No.25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 VLSFLL 11
Db 245 VLSFLL 250
RESULT 12
O6N2_HUMAN STANDARD; PRT; 317 AA.
ID O6N2_HUMAN
AC Q8NGY6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Olfactory receptor 6N2.
GN OR6N2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Suwa M., Sato T., Okouchi I., Arita M., Putani K., Matsumoto S.,
RA Tautumi S., Aburatani H., Asai K., Akiyama Y.;
RT "Genome-wide discovery and analysis of human seven transmembrane helix
RT receptor genes."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Putative odorant receptor.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -!- DATABASE: NAME=Human Olfactory Receptor Data Explorer (HORDE);
CC WWW=<http://bioinformatics.weizmann.ac.il/cgi-bin/HORDE/showgene.pl?key=symbol&>
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB065630; BAC05856.1; -.
DR Genew; HGNC:15035; OR6N2.
DR InterPro; IPR000276; GPCR_Rhodpsn.

```
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 26 46
FT DOMAIN 47 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 75
FT DOMAIN 76 99 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 100 120
FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 160
FT DOMAIN 161 197 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 198 217
FT DOMAIN 218 237 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 238 258
FT DOMAIN 259 271 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 272 292
FT DOMAIN 293 317 CYTOPLASMIC (POTENTIAL).
FT DISULFID 97 189 BY SIMILARITY.
FT CARBOHYD 5 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 317 AA; 35691 MW; D6CAD3D0B0C981D CRC64;

Query Match 30.0%; Score 6; DB 1; Length 317;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFL 10
Db 64 SVLSFL 69

RESULT 13
ID O6K6 HUMAN STANDARD; PRT; 319 AA.
AC Q8NGW6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Olfactory receptor 6K6.
GN OR6K6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA Tautumi S., Aburatani H., Asai K., Akiyama Y.;
RT "Genome-wide discovery and analysis of human seven transmembrane helix
RT receptor genes.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Putative odorant receptor.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -!- DATABASE: NAME=Human Olfactory Receptor Data Explorer (HORDE);
WWW="http://bioinformatics.weizmann.ac.il/cgi-bin/HORDE/showGene.pl?key=symbols&
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB065650; BAC05876.1; -.
Genew; HGNC:15033; OR6K6.
InterPro; IPR000276; GPCR_Rhodopsn.
Pfam; PF00001; 7tm.1; 1.
PRINTS; PR00237; GPCRHHODPSN.
```

```
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 29 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 30 50
FT DOMAIN 51 58 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 59 79
FT DOMAIN 80 103 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 104 124
FT DOMAIN 125 143 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 144 164
FT DOMAIN 165 200 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 201 220
FT DOMAIN 221 240 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 241 261
FT DOMAIN 262 274 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 275 295
FT DOMAIN 296 319 CYTOPLASMIC (POTENTIAL).
FT DISULFID 101 193 BY SIMILARITY.
FT CARBOHYD 9 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 319 AA; 35531 MW; 83A9C35DAAA3B4B9 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 319;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFL 10
Db 68 SVLSFL 73

RESULT 14
ID O5T3 HUMAN STANDARD; PRT; 322 AA.
AC Q8NGG3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Olfactory receptor 5T3.
GN OR5T3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA Tautumi S., Aburatani H., Asai K., Akiyama Y.;
RT "Genome-wide discovery and analysis of human seven transmembrane helix
RT receptor genes.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Putative odorant receptor. Could also be involved in
CC taste perception.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -!- DATABASE: NAME=Human Olfactory Receptor Data Explorer (HORDE);
WWW="http://bioinformatics.weizmann.ac.il/cgi-bin/HORDE/showGene.pl?key=symbols&
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB065837; BAC06056.1; ALT_INIT.
Genew; HGNC:15297; OR5T3.
InterPro; IPR000276; GPCR_Rhodopsn.
Pfam; PF00001; 7tm.1; 1.
PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
```

DR PROSITE; PS00262; G PROTEIN RECEPTOR FL 2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Multigene family; Olfaction.
 FT DOMAIN 1 37 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 38 58 1 (POTENTIAL).
 FT DOMAIN 59 66 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 67 87 2 (POTENTIAL).
 FT DOMAIN 88 111 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 112 132 3 (POTENTIAL).
 FT DOMAIN 133 151 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 152 172 4 (POTENTIAL).
 FT DOMAIN 173 208 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 209 229 5 (POTENTIAL).
 FT DOMAIN 230 249 6 (POTENTIAL).
 FT TRANSMEM 250 270 7 (POTENTIAL).
 FT DOMAIN 271 283 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 284 304 8 (POTENTIAL).
 FT DOMAIN 305 322 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 109 201 BY SIMILARITY.
 FT CARBOHYD 101 17 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 101 17 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 322 AA; 36240 MW; 3BEC6146C34EC6E3 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 322;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 SVLSFL 10
 Db 76 SVLSFL 81

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AY073728; AAL61391.1; -.
 DR EMBL; AY318291; AAP71520.1; -.
 DR InterPro; IPR000276; GPCR_Rhodpsn.
 DR Pfam; PF00001; 7tm 1; 1.
 DR PRINTS; PR00237; GPCR_Rhodopsn.
 DR PROSITE; PS00237; G PROTEIN RECEPTOR FL 1; 1.
 DR PROSITE; PS00262; G PROTEIN RECEPTOR FL 2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Multigene family; Olfaction.
 FT DOMAIN 1 37 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 38 58 1 (POTENTIAL).
 FT DOMAIN 59 66 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 67 87 2 (POTENTIAL).
 FT DOMAIN 88 111 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 112 132 3 (POTENTIAL).
 FT DOMAIN 133 145 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 146 166 4 (POTENTIAL).
 FT DOMAIN 167 208 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 209 229 5 (POTENTIAL).
 FT DOMAIN 230 249 6 (POTENTIAL).
 FT TRANSMEM 250 270 7 (POTENTIAL).
 FT DOMAIN 271 283 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 284 304 7 (POTENTIAL).
 FT DOMAIN 305 330 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 109 201 BY SIMILARITY.
 FT CARBOHYD 17 17 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 330 AA; 37363 MW; 64916C0E6013B97C CRC64;

Query Match 30.0%; Score 6; DB 1; Length 330;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 SVLSFL 10
 Db 76 SVLSFL 81

Search completed: July 19, 2004, 17:13:37
 Job time : 19 secs

RESULT 15
 OR08_MOUSE STANDARD; PRT; 330 AA.
 AC QEVFL3;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Putative olfactory receptor 179-7.
 GN MOR179-7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21676863; PubMed=11802173;
 RA Zhang X., Firestein S.;
 RT "The olfactory receptor gene superfamily of the mouse.";
 RL Nat. Neurosci. 5:124-133(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21664068; PubMed=11875048;
 RA Young J.M., Friedman C., Williams E.M., Ross J.A., Tonnes-Priddy L.,
 RA Trask B.J.;
 RT "Different evolutionary processes shaped the mouse and human
 olfactory receptor gene families.";
 RL Hum. Mol. Genet. 11:535-546(2002).
 RN [3]
 RP ERRATUM.
 RA Young J.M., Friedman C., Williams E.M., Ross J.A., Tonnes-Priddy L.,
 RA Trask B.J.;
 RL Hum. Mol. Genet. 11:1683-1683(2002).
 CC -1- FUNCTION: Potential odorant receptor.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:10:42 ; Search time 36 Seconds
(without alignments)
175.288 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSLVSLFLWTFYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_virus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	7	35.0		355	11	Q9JLY8	Q9jly8 rattus norv
2	7	35.0		355	11	Q8BYW4	Q8byw4 mus musculus
3	7	35.0		355	11	Q8BMH9	Q8bmh9 mus musculus
4	7	35.0		468	17	Q8TUV5	Q8tuv5 methanopyru
5	7	35.0		525	3	O74247	O74247 pichia past
6	7	35.0		1018	16	Q8FWE0	Q8fwe0 corynebacte
7	7	35.0		1019	16	Q8NM51	Q8nm51 corynebacte
8	7	35.0		1369	10	Q8SAZ5	Q8saz5 oryza sativ
9	7	35.0		1865	10	Q81909	Q81909 arabidopsis
10	6	30.0		32	8	Q36494	Q36494 farfantepen
11	6	30.0		35	3	Q874R5	Q874r5 schizosacch
12	6	30.0		50	16	Q989Z9	Q989z9 rhizobium 1
13	6	30.0		57	9	Q858U2	Q858u2 yersinia pe
14	6	30.0		59	8	Q85X72	Q85x72 pinus korai
15	6	30.0		78	12	Q91R80	Q91r80 hibiscus ch
16	6	30.0		87	16	Q9A719	Q9a719 caulobacter

17	6	30.0	93	2	Q8KIN1	Q8kin1 rhizobium e
18	6	30.0	101	3	Q9P7G3	Q9p7g3 schizosacch
19	6	30.0	108	12	Q8B9M9	Q8b9m9 rachiplusia
20	6	30.0	110	12	Q92387	Q92387 bombyx mori
21	6	30.0	111	17	O59222	O59222 pyrococcus
22	6	30.0	112	10	Q9AS01	Q9as01 oryza sativ
23	6	30.0	139	5	Q9N4G1	Q9n4g1 caenorhabdi
24	6	30.0	159	16	Q98KG3	Q98kg3 rhizobium 1
25	6	30.0	165	16	Q8XRZ3	Q8xrz3 ralstonia s
26	6	30.0	171	17	Q8TYN9	Q8ty99 methanopyru
27	6	30.0	180	17	Q97V82	Q97v82 sulfolobus
28	6	30.0	182	12	Q80AH4	Q80ah4 hepatitis b
29	6	30.0	183	16	Q9HXU2	Q9hxu2 pseudomonas
30	6	30.0	192	4	Q8N1Z0	Q8n1z0 homo sapien
31	6	30.0	192	10	Q7XR59	Q7xr59 oryza sativ
32	6	30.0	194	4	O43625	O43625 homo sapien
33	6	30.0	202	17	Q82VF2	Q82vf2 pyrobaculum
34	6	30.0	203	10	Q9AT41	Q9at41 populus tre
35	6	30.0	205	16	Q8X7K5	Q8x7k5 escherichia
36	6	30.0	214	3	Q9UTW4	Q9utw4 schizosacch
37	6	30.0	216	11	Q8C9S7	Q8c9s7 mus musculu
38	6	30.0	217	16	Q8DTV7	Q8dtv7 streptococc
39	6	30.0	223	16	Q8XTY6	Q8xtv6 ralstonia s
40	6	30.0	226	16	O05187	O05187 bacillus su
41	6	30.0	246	16	Q9XDD7	Q9xdd7 brucella me
42	6	30.0	247	17	O26219	O26219 methanobact
43	6	30.0	254	10	Q9FG62	Q9fg62 arabidopsis
44	6	30.0	255	4	Q9H3S2	Q9h3s2 homo sapien
45	6	30.0	255	4	Q7Z6E7	Q7z6e7 homo sapien

ALIGNMENTS

RESULT 1

Q9JLY8 ID Q9JLY8 PRELIMINARY; PRT; 355 AA.
AC Q9JLY8; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Macrophage inflammatory protein-1 alpha receptor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD;
RX MEDLINE=20555330; PubMed=11091494;
RA Waller A., Nayee P., Czaplowski L.G.;
RT "Identification and characterization of a rat macrophage inflammatory protein-1 alpha receptor."
RL J. Hematother. Stem Cell Res. 9:703-710(2000).
DR EMBL: AF119381; AAF34340.1; -
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR InterPro: IPR0001584; P:G-protein coupled receptor protein signalin. .; IEA.
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCR_RHODPSN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 355 AA; 40838 MW; 2FEB8661D1B6E075 CRC64;

Query Match 35.0%; Score 7; DB 11; Length 355;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLWTPY 15
|||||

```
Db          249 FLLWTPY 255

RESULT 2
Q8BVM4
ID Q8BVM4 PRELIMINARY; PRT; 355 AA.
AC Q8BVM4;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Chemokine.
GN CCRI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
  60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK076275; BAC36284.1; -.
DR MGD; MGI:104618; Ccr1.
DR GO; GO:0016493; F-C-C chemokine receptor activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0030595; P:immune cell chemotaxis; IDA.
DR GO; GO:0006954; P:inflammatory response; IMP.
DR GO; GO:0030099; P:myeloid blood cell differentiation; IMP.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm.1.1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1.1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2.1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2.1; 1.
SQ SEQUENCE 355 AA; 40908 MW; F41E2ED5150F7725 CRC64;

Query Match          35.0%; Score 7; DB 11; Length 355;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          9 FLLWTPY 15
          |||||
Db          249 FLLWTPY 255

RESULT 3
Q8BMH9
ID Q8BMH9 PRELIMINARY; PRT; 355 AA.
AC Q8BMH9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Chemokine.
GN CCRI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Forelimb;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
  60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK031109; BAC27256.1; -.
DR MGD; MGI:104618; Ccr1.
DR GO; GO:0016493; F-C-C chemokine receptor activity; IDA.
DR
```

```
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0030595; P:immune cell chemotaxis; IDA.
DR GO; GO:0006954; P:inflammatory response; IMP.
DR GO; GO:0030099; P:myeloid blood cell differentiation; IMP.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm.1.1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1.1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2.1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2.1; 1.
SQ SEQUENCE 355 AA; 40891 MW; E0B4E87FABC4ED CRC64;

Query Match          35.0%; Score 7; DB 11; Length 355;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          9 FLLWTPY 15
          |||||
Db          249 FLLWTPY 255

RESULT 4
Q8TUV5
ID Q8TUV5 PRELIMINARY; PRT; 468 AA.
AC Q8TUV5;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Uncharacterized membrane protein.
GN MK1848.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natile D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlyavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
  and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
DR EMBL; AE010453; AAM02861.1; -.
DR Complete proteome.
SQ SEQUENCE 468 AA; 51789 MW; AFC9C3586D4A4E78 CRC64;

Query Match          35.0%; Score 7; DB 17; Length 468;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY          5 SVLSFLL 11
          |||||
Db          303 SVLSFLL 309

RESULT 5
Q74247
ID Q74247 PRELIMINARY; PRT; 525 AA.
AC Q74247;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Inositol 1-phosphate synthase (EC 5.5.1.4).
GN INOL.
OS Pichia pastoris (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Pichia.
OX NCBI_TaxID=4922;
RN [1]
RP SEQUENCE OF 21-525 FROM N.A.
RX MEDLINE=99392452; PubMed=10463156;
```


RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=O.sativa (japonica cultivar-group);
RA Wang R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,
RA Rambo T., Sasaki C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=O.sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "in-depth view of structure, activity, and evolution of rice
RL chromosome 10";
RN Science 300:1566-1569 (2003).
[4]
RP SEQUENCE FROM N.A.
RC SPECIES=O.sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC093568; AAL78106.1; -
DR EMBL; AC122146; AAM47296.1; -
DR EMBL; AE017065; AAP52471.1; -
DR Gramene; Q8SAZ5; -
DR GO; GO:0005634; C:nucleus; IEA.
DR InterPro; IPR007658; DUF594
DR InterPro; IPR001005; MYB_DNA_binding.
DR Pfam; PF04578; DUF594; 2.
DR PROSITE; PS00037; MYB_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 1369 AA; 155170 MW; 35C051EALBICFC96 CRC64;

Query Match 35.0%; Score 7; DB 10; Length 1369;

Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLW 12
|||||
DB 742 VLSFLW 748

RESULT 9

OB1909 PRELIMINARY; PRT; 1865 AA.
AC OB1909;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE T7123.15 protein.
GN T7123.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Federspiel N.A., Conway A.B., Conway A.R., Davis K., Brendel V.,
RA Palm C.J., Au M., Araujo R., Chung E., Kurtz D.B., Buehler E.,
RA Dewar K., Feng J., Kim C., Li Y., Shinn P., Sun H., Oji O.,
RA Osborne B., Shen Y.K., Toriumi M., Vyotskaia V., Yu G., Theologis A.,
RA Ecker J., Davis R.W.;
RT "Genomic sequence of Arabidopsis BAC T7123.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89959; AAC24379.1; -
DR PIR; G86152; G86152.
DR InterPro; IPR007196; Not1.
DR Pfam; PF04054; Not1; 1.
SQ SEQUENCE 1865 AA; 208122 MW; 23C4D4DFF196EF11 CRC64;

Query Match 35.0%; Score 7; DB 10; Length 1865;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 HSVLSFL 10
|||||
DB 178 HSVLSFL 184

RESULT 10

Q36494 PRELIMINARY; PRT; 32 AA.
AC Q36494;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NADH dehydrogenase subunit 5 (Fragment).
GN NADH5.
OS Farfantepenaeus notialis (Southern pink shrimp).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC Penaeidae; Farfantepenaeus.
OX NCBI_TaxID=44108;
RN [1]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E., Dennebouy N., Oliva-Suarez M., Mounolou J.C.,
RA Monnerot M.;
RT "Partial sequence of the shrimp Penaeus notialis mitochondrial
RT genome";
RL C. R. Acad. Sci., D, Sci. Nat. 319:473-486(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E.;
RL Thesis (1997), Paris-Sud, UFR d'Orsay, France.
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99298297; PubMed=10368442;
RA Garcia-Machado E., Pempera M., Dennebouy N., Oliva-Suarez M.,
RA Mounolou J.C., Monnerot M.;
RT "Mitochondrial genes collectively suggest the paraphyly of Crustacea
RT with respect to Insecta";
RL J. Mol. Evol. 49:142-149(1999).
DR EMBL; X84350; CAB40370.1; -
DR GO; GO:0005739; C:mitochondrion; IEA.
KW Mitochondrion.
FT NON_TER 1 1
FT NON_TER 32 32
SQ SEQUENCE 32 AA; 3783 MW; 32A62974E4937800 CRC64;

Query Match 30.0%; Score 6; DB 8; Length 32;

Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LSFLW 12
|||||
DB 18 LSFLW 23

RESULT 11

Q874R5 PRELIMINARY; PRT; 35 AA.
AC Q874R5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mde5 protein (Fragment).
GN MDE5.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Raylin K., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98944; CAD62442.1; -.
FT NON TER 35
SQ SEQUENCE 35 AA; 4064 MW; 3C800B48D391AC2E CRC64;

Query Match 30.0%; Score 6; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFL 11
Db 13 VLSFL 18

RESULT 12
Q989Z9 PRELIMINARY; PRT; 50 AA.
AC Q989Z9;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Transposase.
GN MSR6219.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003008; BAB52545.1; -.
KW Complete proteome.
SQ SEQUENCE 50 AA; 5204 MW; EFD8E5BD1587B0B CRC64;

Query Match 30.0%; Score 6; DB 16; Length 50;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VLSFL 10
Db 11 VLSFL 16

RESULT 13
Q858J2 PRELIMINARY; PRT; 57 AA.
AC Q858J2;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Protein 19.3.
OS Yersinia pestis phase phiA1122.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=227720;
RN [1]
RP SEQUENCE FROM N.A.
RA Garcia E., Elliott J.M., Ramanculov E., Chain P.S., Chu M.C.,
RA Molineux I.J.;
RT "The genome sequence of Yersinia pestis bacteriophage phiA1122 reveals
RT an intimate history with the coliphage T3.";

```

```

RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY247822; AAP20547.1; -.
SQ SEQUENCE 57 AA; 6418 MW; A0A744FDB0EFB564 CRC64;

Query Match 30.0%; Score 6; DB 9; Length 57;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VLSFL 10
Db 38 VLSFL 43

RESULT 14
Q85X72 PRELIMINARY; PRT; 59 AA.
AC Q85X72;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PSII K protein.
GN PSBK.
OS Pinus koraiensis (Korean pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=88728;
RN [1]
RP SEQUENCE FROM N.A.
RA Noh E.W., Lee J.S., Choi Y.I., Han M.S., Yi Y.S., Han S.U.;
RT "Complete nucleotide sequence of Pinus koraiensis.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY228468; AAO73990.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0009538; C:photosystem I reaction center; IEA.
DR GO; GO:0015979; P:photosynthesis; IEA.
DR InterPro; IPR003687; PSII_PSBK.
DR Pfam; PF02533; PsbK; 1.
KW Chloroplast.
SQ SEQUENCE 59 AA; 6672 MW; 5AB7C7F6D6701FED CRC64;

Query Match 30.0%; Score 6; DB 8; Length 59;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFL 11
Db 43 VLSFL 48

RESULT 15
Q9IR80 PRELIMINARY; PRT; 78 AA.
AC Q9IR80;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Putative movement protein p9.
OS Hibiscus chlorotic ringspot virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC Carmovirus.
OX NCBI_TaxID=53181;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20173705; PubMed=10708431;
RA Huang M., Koh D.C.Y., Weng L., Chang M.L., Yap Y.K., Zhang Y.K.,
RA Wong S.M.;
RT "Complete nucleotide sequence and genome organization of hibiscus
RT chlorotic ringspot virus, a new member of the genus carmovirus:
RT evidence for the presence and expression of two novel open reading
RT frames.";
RL J. Virol. 74:3149-3155(2000).
DR EMBL; X86448; CAB81770.1; -.

```

SQ SEQUENCE 78 AA; 8995 MW; 1E18EB61BB2C5CAB CRC64;
Query Match 30.0%; Score 6; DB 12; Length 78;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 VLSFLL 11
|||
Db 7 VLSFLL 12

Search completed: July 19, 2004, 17:14:20
Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 00:04:45 ; Search time 1705 Seconds
(without alignments)
508.423 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 99
Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO spool_p/US10799747/runat 19072004 161416 21440/app query.fasta_1.199
-DB=GenEmbl -QFWT=fastap -SUPFIX=rg -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747 @CGN 1.1 5600 @runat 19072004 161416 21440 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_nam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	98	99.0	1434	6	BD078422	BD078422 101 human
2	98	99.0	3452	9	BC033650	BC033650 Homo sapi
3	98	99.0	3522	9	HS0804706	AL833393 Homo sapi
C 4	98	99.0	175081	9	AC021151	AC021151 Homo sapi
C 5	98	99.0	281662	2	AC117374	AC117374 Homo sapi
C 6	90	90.9	2020	9	HS0800227	AL049442 Homo sapi
C 7	58	58.6	194420	2	AC124133	AC124133 Mus muscu
C 8	58	58.6	244669	2	AC096099	AC096099 Rattus no
C 9	58	58.6	245830	2	AC140313	AC140313 Mus muscu
C 10	57	57.6	156471	2	AC013568	AC013568 Homo sapi
C 11	57	57.6	192867	9	AL358815	AL358815 Human DNA
C 12	56	56.6	81715	9	AC024680	AC024680 Homo sapi
C 13	56	56.6	165127	9	AC091047	AC091047 Homo sapi
C 14	56	56.6	233717	2	AC140381	AC140381 Mus muscu
C 15	55	55.6	98558	9	AL137138	AL137138 Human DNA
16	54.5	55.1	138187	10	MM0461	AF027865 Mus muscu
17	54	54.5	51559	2	AC099861	AC099861 Mus muscu
18	54	54.5	137930	2	AC141193	AC141193 Rattus no
19	54	54.5	201465	2	AC122424	AC122424 Mus muscu
C 20	54	54.5	226253	2	AC133434	AC133434 Rattus no
C 21	54	54.5	227323	2	AC117350	AC117350 Rattus no
C 22	54	54.5	272007	2	AC110094	AC110094 Rattus no
C 23	53	53.5	30040	2	AC015491	AC015491 Homo sapi
C 24	53	53.5	60464	2	AL359673	AL359673 Homo sapi
C 25	53	53.5	60828	9	HS1191B2	AL022237 Human DNA
C 26	53	53.5	63007	2	AC037446	AC037446 Homo sapi
C 27	53	53.5	63731	2	AL360016.4	Continuation (5 of
C 28	53	53.5	115198	2	AL139223	Continuation (5 of
C 29	53	53.5	138571	2	AC116710	AC116710 Homo sapi
C 30	53	53.5	145111	2	AL390245	AL390245 Homo sapi
C 31	53	53.5	148679	9	AC099788	AC099788 Homo sapi
C 32	53	53.5	148729	9	AC119724	AC119724 Homo sapi
C 33	53	53.5	153615	2	AC111043	AC111043 Mus muscu
C 34	53	53.5	176308	2	AC124005	AC124005 Mus muscu
C 35	53	53.5	176684	2	AC023127	AC023127 Homo sapi
C 36	53	53.5	204495	10	AL591884	AL591884 Mouse DNA
C 37	53	53.5	209513	2	AC122952	AC122952 Rattus no
C 38	53	53.5	227363	2	AL592002	AL592002 Mus muscu
C 39	53	53.5	277861	2	HSAC000406	AC000406 Homo sapi
C 40	52	52.5	4194	9	AK126784	AK126784 Homo sapi
C 41	52	52.5	49743	2	AC100574	AC100574 Mus muscu
C 42	52	52.5	61993	2	AC137891	AC137891 Homo sapi
C 43	52	52.5	62550	2	AC136338	AC136338 Homo sapi
C 44	52	52.5	66251	2	AC115984	AC115984 Homo sapi
45	52	52.5	94308	9	AC005480	AC005480 Homo sapi

ALIGNMENTS

RESULT 1

```

BD078422
LOCUS       BD078422               1434 bp    DNA        linear    PAT 27-AUG-2002
DEFINITION  101 human secretory proteins.
ACCESSION   BD078422
VERSION     BD078422.1 GI:22624025
KEYWORDS    JP 2001519156-A/11.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1434)
AUTHORS     Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
            Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
            Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
            Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
            Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
            Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
            Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
            Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
            McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
            Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
            Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
            Fahey, J., Helton, E., Kettner, M., Madan, A., Young, A.C., Rodrigues, S.,
            Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
            Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
            Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
            Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E.,
            Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
            Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
            Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22398257
MEDLINE     12477932
PUBMED      12477932
2 (bases 1 to 3452)
AUTHORS     Strausberg, R.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
            Email: cgabbs@mail.nih.gov
            Tissue procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc.mgc@nih.gov
            Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
            Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
            Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
            Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
            Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
            McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
            Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
            Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov
Series: IRAK Plate: 69 Row: n Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14249445.

FEATURES             Location/Qualifiers
     source            1..3452
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:44889 IMAGE:5574637"
                     /tissue_type="Duodenum, adenocarcinoma"
                     /clone_lib="NIH MGC 88"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene              1..3452
                     /gene="FLJ14431"
                     /db_xref="LocusID:84869"
     CDS               159..872

ORGANISM             Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3452)
AUTHORS     Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
            Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
            Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
            Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
            Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
            Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
            Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
            Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
            McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
            Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
            Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
            Fahey, J., Helton, E., Kettner, M., Madan, A., Young, A.C., Rodrigues, S.,
            Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
            Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
            Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
            Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E.,
            Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
            Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
            Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22398257
MEDLINE     12477932
PUBMED      12477932
2 (bases 1 to 3452)
AUTHORS     Strausberg, R.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
            Email: cgabbs@mail.nih.gov
            Tissue procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc.mgc@nih.gov
            Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
            Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
            Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
            Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
            Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
            McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
            Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
            Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov
Series: IRAK Plate: 69 Row: n Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14249445.

FEATURES             Location/Qualifiers
     source            1..3452
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:44889 IMAGE:5574637"
                     /tissue_type="Duodenum, adenocarcinoma"
                     /clone_lib="NIH MGC 88"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene              1..3452
                     /gene="FLJ14431"
                     /db_xref="LocusID:84869"
     CDS               159..872

```



```

/codon_start=1
/product="FLJ14431 protein"
/protein_id="AAH33650.1"
/db_xref="GI:21707067"
/db_xref="LocusID:84869"
/translation="MDKVCAGGSGRIGRAVAQLMARKGYRLAVIARNLEGAKAAG
DIGGDHLASCDVAKHDVQNTFELEKHLGRVFNVAAGINRDGLVKTEDMVS
QHTNLGSLMTCVKAMRTMIQQQSGIVNGSVGLKNGSQSVYSASKGLGVFSR
ALAKEVARKIRVNVAPGVFHTDWTDKLKEHLKKNIPLGRFGETIEVAHAVVFLE
SPYITGHVLVDVGLQLIL"
misc_feature
159..857
/note="adh short; Region: short chain dehydrogenase. This
family contains a wide variety of dehydrogenases"
/db_xref="CDD:pfam00106"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 2,11e-07 Length: 3452
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 9 Gaps: 0

```

US-10-799-747-116 (1-20) x BC033650 (1-3452)

```

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTriThrProTyrAlaLeuLysSer 19
|||||
Db 2542 ATGGCAGCCCAATTCAGTCTTGAGTTTCTTCTCGACACCTTATGCTCTGAAATCA 2598
|||||

RESULT 3
HSM804706
LOCUS Homo sapiens mRNA; cDNA DKFp762K109 (from clone DKFp762K109).
ACCESSION AL833393
VERSION AL833393.1 GI:21734029
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3522)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.
JOURNAL Direct Submission
TITLE Submitted (09-JUL-2002) 1, D-85764 Neuberberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Medicinmix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. This clone
(DKFp762K109) is available at the RZPD in Berlin. Please contact
the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.

```

FEATURES

```

source
1..3522
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFp762K109"
/db_xref="taxon:9606"
/clone="DKFp762K109"
/issue_type="melanoma (MeWo cell line)"
/clone_lib="762 (synonym: hmel2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
3420..3425
polya_signal
3420..3425
polya_site
3440

```

ORIGIN

```

Alignment Scores:
Pred. No.: 2,16e-07 Length: 3522
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0

```

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 9 Gaps: 0

```

US-10-799-747-116 (1-20) x HSM804706 (1-3522)

```

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTriThrProTyrAlaLeuLysSer 19
|||||
Db 2552 ATGGCAGCCCAATTCAGTCTTGAGTTTCTTCTCGACACCTTATGCTCTGAAATCA 2608
|||||

RESULT 4
AC021151/c
LOCUS Homo sapiens BAC clone RP11-483A20 from 4, complete sequence.
DEFINITION AC021151
ACCESSION AC021151
VERSION AC021151.8 GI:15145598
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 175081)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE
2 (bases 1 to 175081)
AUTHORS Kyung,K. and Abbott,A.
TITLE The sequence of Homo sapiens BAC clone RP11-483A20
JOURNAL Unpublished (2001)
REFERENCE
3 (bases 1 to 175081)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
4 (bases 1 to 175081)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
5 (bases 1 to 175081)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 9, 2001 this sequence version replaced gi:13877272.

```

COMMENT

```

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplense@wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0483A20

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-36G9. Actual start of this clone is at base position 1 of RP11-483A20; actual end is at base position 175081 of RP11-483A20.

FEATURES

source	Location/Qualifiers
1..175081	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-483A20"
	/clone_lib="RPCI-11"
1..3267	/rpt_family="L1"
3270..3453	/rpt_family="Alu"
3434..3464	/rpt_family="Alu"
	/rpt_family="(A)n"
3455..4161	/rpt_family="L1"
7949..7956	/note="match to EST AW006057 (NID:95854835) wz81d04.x1"
10197..10489	/rpt_family="L1"
10480..10501	/rpt_family="AT_rich"
10495..10665	/rpt_family="L1"
10731..11051	/rpt_family="L1"
11052..11576	/rpt_family="L1"
11564..11598	/rpt_family="L1"
11642..12324	/rpt_family="(A)n"
11769..11794	/rpt_family="L1"
11939..11962	/rpt_family="AT_rich"
12335..12412	/rpt_family="AT_rich"
12392..12621	/rpt_family="L1"
12788..13096	/rpt_family="L1"
13069..13102	/rpt_family="Alu"
13325..13667	/rpt_family="(CAAAA)n"
13444..13484	/rpt_family="Alu"
13640..13686	/rpt_family="(TAAA)n"
14249..14422	/rpt_family="(GAAAA)n"
14980..15182	/rpt_family="MERL_type"
	/rpt_family="L1"
repeat_region	15260..15492 /rpt_family="L1"
repeat_region	15483..15985 /rpt_family="MERL_type"
repeat_region	15986..16105 /rpt_family="L1"
repeat_region	16117..16251 /rpt_family="L1"
repeat_region	16288..16356 /rpt_family="L1"
repeat_region	16373..16545 /rpt_family="L1"
repeat_region	16550..18773 /rpt_family="L1"
repeat_region	18770..20218 /rpt_family="L1"
repeat_region	19228..19346 /rpt_family="T-rich"
repeat_region	20219..20514 /rpt_family="Alu"
repeat_region	20515..22484 /rpt_family="L1"
repeat_region	22584..22882 /rpt_family="Alu"
repeat_region	22748..22768 /rpt_family="AT_rich"
repeat_region	22883..23460 /rpt_family="L1"
repeat_region	23461..23747 /rpt_family="MaLR"
repeat_region	23749..24514 /rpt_family="L1"
misc_feature	24456 /note="match to EST AW902102 (NID:98066307)"
misc_feature	24488..24797 /note="match to EST AW902102 (NID:98066307)"
repeat_region	24576..24625 /rpt_family="Mariner"
repeat_region	25484..25787 /rpt_family="Alu"
repeat_region	25758..25787 /rpt_family="(A)n"
repeat_region	25796..26014 /rpt_family="MIR"
repeat_region	26657..26889 /rpt_family="L1"
repeat_region	26890..27205 /rpt_family="MaLR"
repeat_region	27206..27439 /rpt_family="Alu"
repeat_region	27440..27490 /rpt_family="MaLR"
repeat_region	27491..28717 /rpt_family="L1"
repeat_region	28718..28742 /rpt_family="(T)n"
misc_feature	28735..28771 /note="match to EST AA954400 (NID:g3118095) 0001f01.s1"
misc_feature	28747..28826 /note="match to EST AW006057 (NID:95854835) wz81d04.x1"
repeat_region	29022..29308 /rpt_family="L1"
misc_feature	29384..30105 /note="match to EST BF680971 (NID:g11954866)"
misc_feature	29687..30351 /note="match to EST BG429938 (NID:g13336444)"

Alignment Scores:

Pred. No.:	1.81e-05	Length:	175081
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0

DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x AC0211151 (1-175081)

QY 1 MetaAlaHisSerValLeuSerPheLeuLeuTrrPheProTyrAlaLeuLysSer 19
|||||

Db 57641 ATGGCAGCCATTCAGTCTTGAGTTTCTCTCTGACACCTTATGCTCTGAAATCA 57585

RESULT 5
AC117374/c

LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
PROGRESS ***, 48 unordered pieces.

ACCESSION AC117374 AC011694

VERSION AC117374.1 GI:20127352

KEYWORDS HTG; HTGS PHASE1.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 281662)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douchwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gortell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holliday,C., Hollins,B.,
Homi,F., Howard,S., Huber,J., Hulyk,J., Hume,J., Jackson,L.B.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozad,O.R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,B., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 281662)

Worley,K.C.

Direct Submission

Submitted (10-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 281662)

Worley,K.C.

Direct Submission

Submitted (02-MAY-2002) Human Genome Sequencing Center, Department

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:10045385.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCN2
Center clone name: RP11-19D19
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 273063 bases at least Q40
Consensus quality: 294336 bases at least Q30
Consensus quality: 306942 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2464: contig of 2464 bp in length
* 2465 2564: gap of unknown length
* 2565 4603: contig of 2039 bp in length
* 4604 4703: gap of unknown length
* 4704 7113: contig of 2410 bp in length
* 7114 7213: gap of unknown length
* 7214 9517: contig of 2304 bp in length
* 9518 9617: gap of unknown length
* 9618 12107: contig of 2490 bp in length
* 12108 12207: gap of unknown length
* 12208 15014: contig of 2807 bp in length
* 15015 15114: gap of unknown length
* 15115 17331: contig of 2717 bp in length
* 17332 17931: gap of unknown length
* 17932 20580: contig of 2549 bp in length
* 20481 20580: gap of unknown length
* 20581 22700: contig of 2120 bp in length
* 22701 22800: gap of unknown length
* 22801 24909: contig of 2109 bp in length
* 24910 25009: gap of unknown length
* 25010 27313: contig of 2304 bp in length
* 27314 27413: gap of unknown length
* 27414 29798: contig of 2385 bp in length
* 29799 33466: gap of unknown length
* 33467 33666: gap of unknown length
* 33667 36779: contig of 3113 bp in length
* 36680 40472: contig of 3693 bp in length
* 40473 40572: gap of unknown length
* 40573 42684: contig of 2111 bp in length
* 42684 42783: gap of unknown length
* 42784 46319: contig of 3536 bp in length
* 46320 46419: gap of unknown length
* 46420 49507: contig of 3088 bp in length
* 49508 49607: gap of unknown length
* 49608 53212: contig of 3605 bp in length
* 53213 53312: gap of unknown length
* 53313 57267: contig of 3955 bp in length
* 57268 60444: contig of 3077 bp in length
* 60445 60544: gap of unknown length
* 60545 65859: contig of 5315 bp in length
* 65860 65959: gap of unknown length

```

* 65960 69038: contig of 3079 bp in length
* 69039 69138: gap of unknown length
* 69139 73445: contig of 4307 bp in length
* 73446 73545: gap of unknown length
* 73546 78107: contig of 4562 bp in length
* 78108 78207: gap of unknown length
* 78208 82504: contig of 4297 bp in length
* 82505 82604: gap of unknown length
* 82605 86976: contig of 4372 bp in length
* 86977 91315: gap of unknown length
* 91316 91415: contig of 4239 bp in length
* 91416 95084: contig of 3669 bp in length
* 95085 95184: gap of unknown length
* 95185 102348: contig of 7164 bp in length
* 102349 102448: gap of unknown length
* 102449 105973: contig of 3525 bp in length
* 105974 106073: gap of unknown length
* 106074 112891: contig of 6818 bp in length
* 112892 112991: gap of unknown length
* 112992 116221: contig of 3230 bp in length
* 116222 116321: gap of unknown length
* 116322 119798: contig of 3477 bp in length
* 119799 119898: gap of unknown length
* 119900 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140737: gap of unknown length
* 140738 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155057: contig of 9416 bp in length
* 155058 155157: gap of unknown length
* 155158 164357: contig of 9200 bp in length
* 164358 164358: gap of unknown length
* 164359 174832: contig of 10435 bp in length
* 174833 174932: gap of unknown length
* 174933 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 216436: gap of unknown length
* 216437 216537: contig of 17951 bp in length
* 216538 22916: gap of unknown length
* 22917 233016: contig of 16380 bp in length
* 233017 255437: contig of 22421 bp in length
* 255438 255538: gap of unknown length
* 255539 281662: contig of 26125 bp in length.

```

FEATURES

source

```

1. .281662
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-19D19"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 3.1e-05 Length: 281662
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 2 Gaps: 0

```

US-10-799-747-116 (1-20) x AC117374 (1-281662)

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuYsSer 19

```

|||||
Db 145718 ATGGCAGCCCAATTCAGCTTGGATTTCTTCTCTGGACCATATGCTCTGAAATCA 145662
|||||

```

RESULT 6

HSM800227

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

HSM800227 2020 bp mRNA linear PRI 18-FEB-2000
 Homo sapiens mRNA; CDNA DKFZp586N1720 (from clone DKFZp586N1720).
 AL049442
 AL049442.1 GI:4500222
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2020)
 Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
 Direct Submission
 Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152
 Martinsried, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 This clone (DKFZp586N1720) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES

source

```

1. .2020
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp586N1720"
/db_xref="taxon:9606"
/clone="DKFZp586N1720"
/tissue_type="uterus"
/clone_lib="586 (synonym: hutel). Vector pSport1; host  

DH10B; sites NotI + SalI/MluI"
/dev_stage="adult"
1983..1988
polyA_signal
polyA_site
2003

```

ORIGIN

```

Alignment Scores:
Pred. No.: 3.26e-06 Length: 2020
Score: 90.00 Matches: 18
Percent Similarity: 94.74% Conservative: 0
Best Local Similarity: 94.74% Mismatches: 1
Query Match: 90.91% Indels: 0
DB: 9 Gaps: 0

```

US-10-799-747-116 (1-20) x HSM800227 (1-2020)

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuYsSer 19

```

|||||
Db 1115 ATGGCAGCCCAATTCAGCTTGGATTTCTTCTCTGGACCATATGCTCTGAAATCA 1171
|||||

```

RESULT 7

AC124133/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC124133 194420 bp DNA linear HTG 18-DEC-2003
 Mus musculus chromosome 15 clone RP23-218L23 map 15, *** SEQUENCING
 IN PROGRESS ***, 2 ordered pieces.

AC124133
 AC124133.10 GI:40018787
 HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 Mus musculus (house mouse)

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 194420)

Birren,B., Nusbaum,C. and Lander,E.
 Mus musculus chromosome 15, clone RP23-218L23
 Unpublished
 2 (bases 1 to 194420)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,